CLASSIFICATION AND REGRESSION TREES Textbook

Classification and regression trees (CART) are powerful and versatile tools in the realm of machine learning and data analysis. They serve as foundational algorithms for both classification tasks?where the goal is to assign data points to predefined categories?and regression tasks, which involve predicting continuous outcomes. Their intuitive structure, ease of interpretation, and ability to handle complex datasets have made them popular among data scientists, statisticians, and domain experts alike. This article explores the fundamental concepts behind classification and regression trees, their construction, advantages, limitations, and practical applications.

Understanding Decision Trees: The Basics

Decision trees are a type of supervised learning algorithm that models decisions and their possible consequences in a tree-like structure. Each internal node represents a decision based on a feature (or attribute), each branch signifies the outcome of the decision, and each leaf node indicates a final output?either a class label or a continuous value.

How Decision Trees Work

The core idea of a decision tree involves recursively splitting the dataset into subsets based on feature values, with the goal of increasing the homogeneity of the resulting subsets. This process continues until a stopping criterion is met, such as a maximum depth or minimum number of samples per leaf.

The steps involved in building a decision tree include:

- Selecting the best feature and threshold to split the data at each node.
- Partitioning data according to this split.
- Recursively repeating the process for each subset.
- Assigning a class label or value to each leaf based on the majority class or average of the subset.

Classification Trees

Classification trees are designed specifically to categorize data points into discrete classes. Their structure and splitting criteria are optimized to maximize the purity of each node, ensuring that data

points within a node belong largely to a single class.

Splitting Criteria for Classification

The effectiveness of a classification tree depends on how well it partitions the data. Common impurity measures used to determine the best split include:

- **Gini Impurity:** Measures the probability of misclassifying a randomly chosen element if it was labeled according to the class distribution in the node. The goal is to minimize Gini impurity at each split.
- **Information Gain (Entropy):** Based on information theory, it quantifies the reduction in entropy after a split. The split with the highest information gain is preferred.

Constructing a Classification Tree

The process involves:

- Calculating the impurity measure for all potential splits across features.
- Selecting the split that results in the greatest reduction of impurity.
- Repeating this process recursively for each child node.
- Assigning class labels to leaves based on the majority class within that node.

Regression Trees

Regression trees extend the decision tree approach to predict continuous, numerical outcomes. Instead of class labels, leaves contain predicted numerical values, often the mean of the target variable within the node.

Splitting Criteria for Regression

The splitting process aims to minimize the variance within each node, leading to more homogeneous groups. Common criteria include:

- **Least Squared Error (LSE):** The split that minimizes the sum of squared residuals (differences between observed and predicted values) is selected.

Constructing a Regression Tree

The steps involve:

- Evaluating potential splits based on the reduction in variance or mean squared error.
- Selecting the split that offers the most significant decrease.
- Recursively applying the process to each child node.
- Assigning a numerical prediction to each leaf, typically the mean value of the target variable in that node.

Advantages of Classification and Regression Trees

Decision trees offer numerous benefits that contribute to their widespread adoption:

- **Interpretability:** The tree structure is easy to understand and visualize, making it accessible even for non-experts.
- **Handling of Different Data Types:** Capable of managing both numerical and categorical data without extensive preprocessing.
- **Non-Parametric Nature:** Do not assume any specific data distribution, allowing flexibility in modeling complex relationships.
- **Feature Selection:** Implicitly perform feature selection by choosing the most informative features at each split.
- **Fast Training and Prediction:** Relatively quick to train and make predictions, especially with optimized implementations.

Limitations and Challenges

Despite their strengths, decision trees also have notable limitations:

- **Overfitting:** Deep trees can model noise in the training data, leading to poor generalization.
- **Instability:** Small variations in data can result in different tree structures.
- **Bias-Variance Tradeoff:** Trees tend to have low bias but high variance, necessitating techniques like pruning or ensemble methods.
- **Greedy Algorithm:** The splitting process is greedy and may not always find the globally optimal tree.

Improving Decision Tree Performance

To address the limitations of a single decision tree, several strategies and ensemble methods can be employed:

Pruning

Pruning involves trimming sections of the tree that do not provide power in predicting outcomes, reducing overfitting. Techniques include:

- Pre-pruning: Stopping the tree growth early based on criteria like maximum depth.
- Post-pruning: Removing branches after the tree is fully grown, often based on validation data.

Ensemble Methods

Combining multiple trees creates more robust models:

- **Random Forests:** Build numerous trees using random subsets of data and features, then aggregate their predictions.
- **Gradient Boosting Machines:** Sequentially add trees that correct the errors of previous ones, leading to highly accurate models.

Practical Applications of Classification and Regression Trees

Decision trees are employed across various domains, including:

- **Medical Diagnosis:** Classifying patients based on symptoms and test results.
- **Financial Analysis:** Credit scoring and risk assessment.
- **Marketing:** Customer segmentation and targeting.
- **Manufacturing:** Predicting equipment failure or quality control issues.
- **Environmental Science:** Modeling ecological phenomena or predicting pollution levels.

Conclusion

Classification and regression trees are versatile, interpretable, and effective algorithms for predictive modeling. Their ability to handle diverse data types and provide transparent decision rules makes them invaluable tools in data science. While they have limitations such as overfitting and instability, these issues can be mitigated through techniques like pruning and ensemble methods. Whether used individually or as part of more complex models like Random Forests or Gradient Boosting Machines, decision trees continue to play a crucial role in tackling real-world problems across industries and disciplines. Embracing their strengths and understanding their limitations allows data practitioners to leverage decision trees effectively and develop robust, accurate predictive models.

Classification and Regression Trees: Unlocking the Power of Decision-Making in Data Science

In the rapidly evolving landscape of data science and machine learning, understanding how to extract meaningful insights from complex datasets is paramount. Among the arsenal of algorithms available, classification and regression trees stand out as intuitive yet powerful tools that bridge the gap between raw data and actionable knowledge. These models, often referred to collectively as decision trees, are prized for their interpretability, flexibility, and robustness in handling various types of data. This article delves into the core concepts, mechanisms, advantages, and practical applications of classification and regression trees, equipping readers with a comprehensive understanding of these essential techniques.

What Are Classification and Regression Trees?

Classification and regression trees (CART) are tree-structured algorithms used for predictive modeling. They operate by recursively partitioning data into subsets based on feature values, ultimately leading to a decision or prediction at the terminal nodes, often called leaves.

- Classification trees are used when the target variable is categorical, such as predicting whether an email is spam or not.
- Regression trees are employed when the target variable is continuous, like estimating house prices or temperature.

Both types of trees share a similar structure and process, differing primarily in their splitting criteria and output.

The Anatomy of a Decision Tree

A decision tree resembles a flowchart, with internal nodes representing tests on features, branches corresponding to outcomes of these tests, and leaves indicating the final prediction.

Key components include:

- Root node: The topmost node representing the entire dataset.
- Internal nodes: Decision points based on feature thresholds.
- Branches: Outcomes of decisions leading to subsequent nodes.
- Leaves: Final predictions?class labels or numerical values.

This hierarchical structure allows for a straightforward visualization of decision rules, making the models inherently interpretable.

How Do Decision Trees Work?

Building the Tree

The process begins with the entire dataset at the root node. The algorithm searches for the feature and threshold that best splits the data into more homogeneous groups regarding the target variable. This splitting continues recursively, creating a tree that grows deeper until stopping criteria are met, such as maximum depth or minimum number of samples in a node.

Splitting Criteria

- Classification trees: Use measures like Gini impurity or entropy (information gain) to evaluate the quality of splits.
- Regression trees: Use variance reduction or mean squared error (MSE) reduction to assess split effectiveness.

The goal is to maximize the purity of resulting nodes, meaning each leaf contains data points that are as similar as possible concerning the target.

Making Predictions

- Classification: Assigns the most common class label within a leaf.
- Regression: Computes the average value of the target variable within a leaf.

Advantages of Using Decision Trees

- Interpretability: Decision trees provide clear, visual rules that humans can understand and trust.
- Handling of Different Data Types: Capable of working with both numerical and categorical variables without extensive preprocessing.
- Non-Linear Relationships: Effectively capture complex, non-linear interactions between features.
- Minimal Data Preparation: Require less data cleaning compared to other models.
- Feature Selection: Implicitly perform feature selection during the split process.

Limitations and Challenges

Despite their strengths, decision trees are not without drawbacks:

- Overfitting: Trees can become overly complex, capturing noise instead of the true underlying pattern.
- Instability: Small changes in data can lead to different trees, affecting consistency.
- Bias: Tend to favor features with more levels or unique values.
- Limited Generalization: Single trees may underperform compared to ensemble methods.

These challenges often motivate the use of ensemble techniques, such as random forests and gradient boosting, which combine multiple trees to improve accuracy and stability.

Pruning and Hyperparameter Tuning: Enhancing Tree Performance

To prevent overfitting and improve the generalization ability of decision trees, various techniques are employed:

- Pruning: Simplifies the tree after it?s built by removing branches that have little predictive power.
- Max Depth: Limits the maximum depth of the tree.
- Minimum Samples per Leaf: Sets the smallest number of samples required to create a leaf.
- Split Criterion Thresholds: Fine-tunes the thresholds for feature splits.

Proper tuning of these hyperparameters ensures a balance between capturing data complexity and maintaining model simplicity.

Practical Applications of Classification and Regression Trees

Decision trees are versatile and find applications across numerous domains:

- Healthcare: Diagnosing diseases based on symptoms and test results.
- Finance: Credit scoring and risk assessment.
- Marketing: Customer segmentation and churn prediction.
- Environmental Science: Modeling climate variables and predicting pollution levels.
- Manufacturing: Fault detection and quality control.

Their interpretability makes them especially valuable in fields where understanding decision logic is crucial.

From Trees to Forests: Ensemble Methods

While individual decision trees are useful, they often serve as the foundation for more sophisticated ensemble methods:

- Random Forests: Aggregate predictions from multiple uncorrelated trees to improve accuracy and reduce overfitting.
- Gradient Boosting Machines: Sequentially build trees that correct the errors of previous trees, leading to highly predictive models.

These ensemble techniques leverage the strengths of decision trees while mitigating their weaknesses, becoming the backbone of many state-of-the-art machine learning solutions.

Final Thoughts

Classification and regression trees embody a blend of simplicity and power that continues to resonate in the data science community. Their intuitive structure allows for transparent decision-making processes, making them accessible to both technical experts and non-specialists. Whether used standalone or as part of ensemble models, decision trees are invaluable tools for extracting insights, making predictions, and informing strategic decisions across industries.

As data complexity grows and the demand for interpretable models increases, the relevance of classification and regression trees is poised to endure. Their ability to adapt to various data types, handle non-linear relationships, and provide clear decision rules ensures they remain a foundational element in the evolving toolkit of machine learning practitioners.

QuestionAnswer What are classification and regression trees (CART), and how are they used in machine learning? CART are decision tree algorithms used for classification (predicting categorical labels) and regression (predicting continuous values). They split data based on feature values to create interpretable models that make predictions by traversing decision paths from root to leaf nodes. How does the CART algorithm determine the best split at each node? CART uses criteria like Gini impurity for classification and mean squared error for regression to evaluate potential splits. It selects the feature and threshold that result in the most significant reduction in impurity or error, optimizing the homogeneity of resulting nodes. What are some advantages of using classification and regression trees? CART models are simple to understand and interpret, handle both numerical and categorical data, require little data preprocessing, and can capture complex, non-linear relationships without requiring feature scaling. What are common methods to prevent overfitting in CART models? Techniques include pruning the tree to remove branches that do not provide significant predictive power, setting a maximum depth, requiring a minimum number of samples per leaf, and using ensemble methods like random forests or gradient boosting to improve generalization. How do ensemble methods improve the performance of CART models? Ensemble methods combine multiple decision trees?such as in random forests or gradient boosting?to reduce variance and bias, leading to more accurate and robust predictions compared to a single tree. What are some limitations of classification and regression trees? CART models can be prone to overfitting, sensitive to small data variations, and may produce unstable trees. They also tend to perform poorly on complex problems unless combined with ensemble techniques. How does feature

importance work in CART models? Feature importance in CART is typically measured by the total reduction in impurity (like Gini impurity or variance) attributable to each feature across all splits in the tree. Features with higher importance contribute more to the model's predictions. Can CART handle multi-class classification problems? Yes, CART can handle multi-class classification by splitting nodes to maximize the purity for multiple classes, often using criteria like Gini impurity or entropy for multi-class splits.

Related keywords: decision trees, supervised learning, machine learning, predictive modeling, CART, data mining, recursive partitioning, feature selection, overfitting, model interpretability

Algebra 2 regents regression analysis

Algebra 2 Regents Regression Analysis is a fundamental skill for students preparing for the New York State Regents exam and for those seeking a deeper understanding of statistical relationships in algebra. Regression analysis helps students interpret data, identify trends, and make predictions based on mathematical models. Mastering this topic is essential for achieving success on the exam and for applying algebraic concepts to real-world problems.

Understanding Regression Analysis in Algebra 2

Regression analysis is a statistical method used to examine the relationship between a dependent variable and one or more independent variables. In Algebra 2, students primarily focus on linear regression, which models data with a straight line, although quadratic and exponential regressions are also explored in advanced contexts. The goal is to find the best-fit line or curve that describes the data and allows for predictions.

What Is Regression Analysis?

Regression analysis involves fitting a mathematical model to a set of data points. The most common form is linear regression, which seeks to find the line of best fit through the data, minimizing the sum of the squared differences (residuals) between observed and predicted values.

Why Is Regression Analysis Important?

- It helps interpret relationships between variables.
- It allows for making predictions based on data trends.
- It enhances problem-solving skills in algebra and statistics.

- It is a key component of the Algebra 2 Regents exam.

Key Concepts in Regression Analysis for Algebra 2

Understanding the core concepts is vital for mastering regression analysis. Here are the main ideas you need to grasp:

Linear Regression

Linear regression models the relationship between two variables with a straight line, described by the equation:

$$ y = mx + b $$

where:

- y is the dependent variable,
- x is the independent variable,
- m is the slope,
- b is the y-intercept.

Scatter Plots

A scatter plot visually displays data points to identify the nature of the relationship:

- Positive correlation: as x increases, y increases.
- Negative correlation: as x increases, y decreases.
- No correlation: no discernible pattern.

Line of Best Fit

The line that best represents the data, often found using least squares regression. It minimizes the total squared residuals and provides the equation for prediction.

Correlation Coefficient (r)

A numerical measure of the strength and direction of the linear relationship:

- \(r \) close to 1 indicates a strong positive correlation.
- \(r \) close to -1 indicates a strong negative correlation.
- \(r \) near 0 indicates no correlation.

Coefficient of Determination (R^2)

Represents the proportion of variance in the dependent variable explained by the independent variable:

- Ranges from 0 to 1.
- Higher R^2 indicates a better fit.

Performing Regression Analysis on the Algebra 2 Regents

Students should understand the step-by-step process to perform regression analysis during the exam:

1. Plot the Data

Create a scatter plot to visualize the relationship between variables.

2. Find the Line of Best Fit

Use a calculator or graphing tool to determine the regression line:

- In a graphing calculator, access the regression feature.
- In real-world scenarios, use software like Desmos or Excel.

3. Write the Equation

Record the regression equation in the form:

\[y = mx + b \]

or appropriate for quadratic or exponential models.

4. Interpret the Results

- Analyze the slope to understand the rate of change.
- Use the y-intercept for context.
- Evaluate \(r \) and R^2 for the strength of the model.

5. Make Predictions

Use the equation to estimate \(y \) for given \(x \) values.

Common Types of Regression in Algebra 2

While linear regression is the most common, students may encounter other types:

Quadratic Regression

Models data with a parabola:

\[y = ax^2 + bx + c \]

Useful when data shows a curved trend.

Exponential Regression

Models exponential growth or decay:

\[y = a \times b^x \]

Common in contexts like population growth or radioactive decay.

Logarithmic Regression

Useful for data that increases rapidly and then levels off:

\[y = a \times \log(x) + b \]

Applying Regression Analysis to Real-World Problems

Regression analysis is not just an academic skill; it has practical applications:

- Predicting sales based on advertising expenditure.
- Estimating population growth over time.

- Analyzing the relationship between temperature and ice cream sales.
- Assessing the impact of study time on test scores.

In the exam, you might be asked to interpret regression results or use the model to make predictions, emphasizing the importance of understanding both the calculations and their real-world implications.

Tips for Success on the Algebra 2 Regents Regression Problems

To excel in regression analysis questions, consider these strategies:

- Familiarize yourself with graphing calculator functions for regression analysis.
- Practice plotting data points and interpreting scatter plots.
- Always write the regression equation clearly and check the slope and intercept for reasonableness.
- Understand how to interpret \(r \) and R^2 values to assess model accuracy.
- Be prepared to explain what the regression line indicates about the data relationship.
- Work through multiple practice problems to build confidence and speed.

Resources for Learning Regression Analysis

Students looking to improve their regression analysis skills can utilize various resources:

- Graphing calculator tutorials (Texas Instruments, Casio)
- Online graphing tools like Desmos
- Practice worksheets aligned with Algebra 2 Regents standards
- Video tutorials on regression concepts
- Study groups and tutoring sessions focusing on data analysis

Conclusion

Mastering **algebra 2 regents regression analysis** is crucial for success on the New York State Regents exam and for developing a solid understanding of data relationships. By understanding the core concepts, practicing the steps to perform regression, and applying these skills to real-world scenarios, students can confidently interpret data, create models, and make informed predictions. Regular practice, the use of technological tools, and a clear grasp of the underlying principles will ensure readiness to tackle regression analysis questions effectively and achieve a high score on the exam.

Algebra 2 Regents Regression Analysis: A Comprehensive Investigation

Regression analysis is a foundational statistical tool in Algebra 2, integral to understanding relationships between variables and predicting future outcomes. The Algebra 2 Regents exam, a critical assessment for high school students in New York State, emphasizes mastery of regression concepts, including linear and nonlinear models. This article delves into the intricacies of regression analysis as presented in the Algebra 2 Regents, exploring its theoretical underpinnings, practical applications, common pitfalls, and instructional strategies to enhance understanding and performance.

Understanding Regression Analysis in Algebra 2

Regression analysis involves identifying and quantifying the relationship between a dependent variable and one or more independent variables. In the context of the Algebra 2 Regents, students primarily focus on linear regression, though quadratic and exponential regressions are also pertinent.

Theoretical Foundations

At its core, regression analysis seeks to fit a model that best describes the data. The most common type, linear regression, assumes a straight-line relationship:

$$ y = ax + b $$

where:

- y is the dependent variable,
- x is the independent variable,
- a is the slope, representing the rate of change,
- b is the y-intercept, indicating the starting point.

The goal is to find the line that minimizes the sum of squared residuals?the differences between observed and predicted values.

Key Concepts and Terminology

- Correlation coefficient (r): Measures the strength and direction of the linear relationship between variables, ranging from -1 to 1.
- Line of best fit: The regression line that minimizes the sum of squared residuals.

- Residuals: The difference between observed \(y \) values and those predicted by the regression line.
- Coefficient of determination (r^2): Indicates the proportion of variance in the dependent variable explained by the independent variable.

Regression Analysis on the Algebra 2 Regents Exam

The Regents exam assesses students? ability to interpret, compute, and apply regression analysis in various contexts.

Types of Regression Covered

- Linear Regression: Most commonly tested; students must interpret scatterplots, compute lines of best fit, and analyze residual plots.
- Quadratic and Exponential Regression: Less frequently, but students should understand how to identify and interpret these models, especially in context-based problems.

Common Question Formats

- Interpreting Regression Equations: Understanding what the slope and intercept signify in real-world context.
- Analyzing Scatterplots: Determining whether a linear model is appropriate based on data patterns.
- Calculating Regression Lines: Using calculator functions or formulae to find the best-fit line.
- Evaluating Model Fit: Using \(r \), \(r^2 \), and residual plots to assess the quality of the regression.
- Prediction and Extrapolation: Making predictions within and beyond the data range, with appropriate caution.

Methodologies and Tools for Regression Analysis

Students are expected to utilize both graphical and algebraic methods.

Calculator Usage

Graphing calculators (such as TI-83/84 series) are essential for regression tasks:

- Input data into lists.

- Use the ?LinReg? function (e.g., `LinReg(ax+b)`) to find the line of best fit.
- Interpret the output coefficients.
- Generate residual plots to assess fit.

Manual Computation and Interpretation

While calculator use is standard, understanding the underlying formulas enhances conceptual clarity:

- Computing the mean of x and y.
- Calculating sums of squares and cross-products.
- Deriving the slope and intercept through the least squares formulas.

Common Challenges and Misconceptions

Despite the straightforward nature of regression analysis, students often encounter pitfalls:

Misinterpreting Correlation

A common misconception is equating correlation with causation. While a high r indicates a strong linear relationship, it does not imply that one variable causes changes in the other.

Extrapolation Risks

Using the regression line to predict outside the data range can be misleading. The model may not hold beyond observed values.

Overfitting and Underfitting Data

Choosing an overly complex model (e.g., quadratic when a linear model suffices) or oversimplifying can lead to inaccurate conclusions.

Ignoring Residual Patterns

Residual plots should display random scatter. Patterns suggest the model may not be appropriate.

Instructional Strategies for Mastery

Effective teaching of regression analysis in Algebra 2 involves integrating conceptual understanding with practical application.

Visual Learning

- Use scatterplots to illustrate data patterns.
- Demonstrate how the line of best fit minimizes residuals.
- Show residual plots to assess fit quality.

Hands-On Practice

- Assign data collection projects (e.g., tracking students? study hours vs. test scores).
- Use calculator simulations for regression computations.
- Incorporate real-world scenarios such as economics, biology, or physics.

Critical Thinking and Interpretation

- Pose questions that require students to interpret regression equations contextually.
- Discuss the limitations of models and the importance of residual analysis.

Conclusion: The Significance of Regression Analysis in Algebra 2 and Beyond

Regression analysis is not only a key component of the Algebra 2 Regents but also a vital skill for scientific inquiry, data analysis, and decision-making in various fields. Mastery involves understanding the mathematical foundations, interpreting real-world data, and recognizing the limitations of models. As students progress, these skills become indispensable tools for analyzing complex data sets, making informed predictions, and critically evaluating statistical claims.

By thoroughly understanding regression concepts, practicing computational techniques, and engaging in meaningful interpretation, students can excel on the Regents exam and develop a strong foundation for future quantitative reasoning. Educators are encouraged to emphasize conceptual clarity, contextual understanding, and critical analysis to prepare students effectively for both the exam and real-world applications.

In summary, algebra 2 regression analysis serves as a bridge between abstract mathematics and tangible data-driven insights. Its mastery is essential for academic success and for cultivating a data literate mindset necessary in today?s information-rich world.

QuestionAnswer What is regression analysis in Algebra 2 Regents exams? Regression analysis is a statistical method used to model and analyze the relationship between a dependent variable

and one or more independent variables, often through the equation of a line or curve, to make predictions or understand data trends. How do you determine the line of best fit in regression analysis? The line of best fit, often called the least squares regression line, is determined by minimizing the sum of the squared differences between observed and predicted values, which is typically calculated using regression formulas or graphing calculators. What is the meaning of the slope in a regression line for Algebra 2? The slope represents the rate of change of the dependent variable with respect to the independent variable; it indicates how much the dependent variable is expected to increase or decrease as the independent variable increases by one unit. How do you interpret the y-intercept in a regression equation? The y-intercept is the predicted value of the dependent variable when the independent variable is zero; it provides a starting point for the regression line on the graph. What is the purpose of calculating the correlation coefficient in regression analysis? The correlation coefficient measures the strength and direction of the linear relationship between variables, with values close to 1 or -1 indicating a strong positive or negative linear relationship, respectively. How can residuals be used to assess the fit of a regression model? Residuals are the differences between observed and predicted values; analyzing residual plots helps determine if the regression line appropriately models the data or if there are patterns indicating a poor fit. What are common mistakes to avoid when performing regression analysis on the Algebra 2 Regents? Common mistakes include confusing the independent and dependent variables, using the wrong formula for the line of best fit, interpreting the regression line incorrectly, and ignoring residual analysis to check the model's fit. How do you use regression analysis to make predictions on the Algebra 2 Regents? Once the regression equation is found, substitute the value of the independent variable into the equation to predict the value of the dependent variable. What is the significance of the coefficient of determination (R^2) in regression analysis? R^2 indicates the proportion of variance in the dependent variable that can be explained by the independent variable(s); values closer to 1 mean a better fit. Are there specific types of regression analysis emphasized in the Algebra 2 Regents exam? Yes, simple linear regression (line of best fit) is most commonly emphasized, but students may also encounter questions involving quadratic or exponential models depending on the data context.

Related keywords: algebra 2 regents, regression analysis, linear regression, quadratic regression, statistical analysis, regression equation, graphing regression, least squares method, correlation coefficient, regression problems

Read the full article online: Algebra 2 regents regression analysis