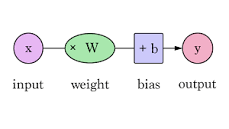
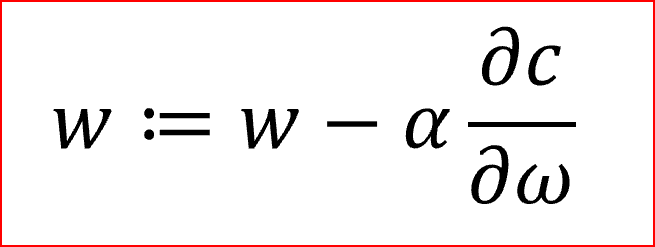
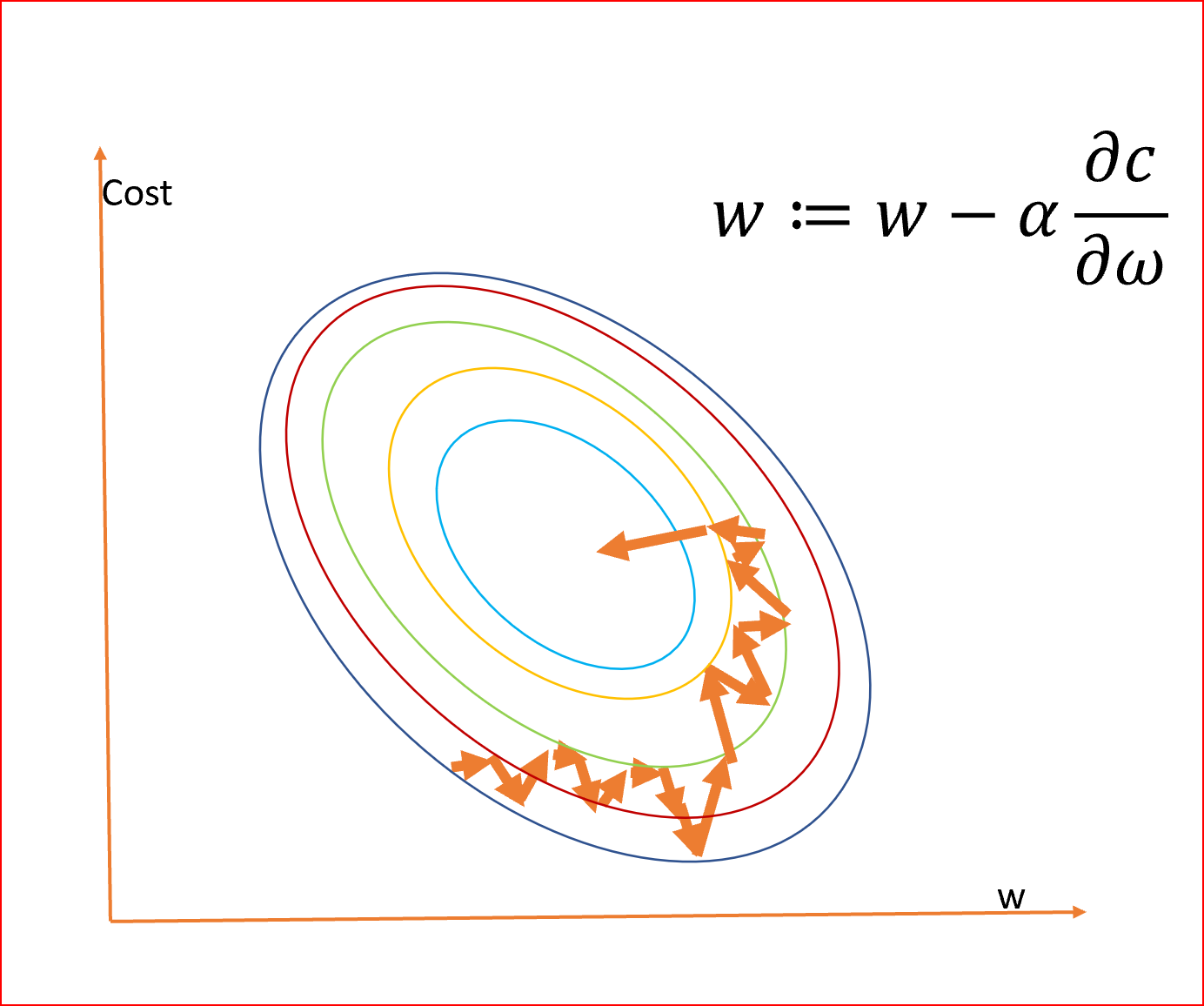
* Linear regression
  + Supervised algorithm
  + Advantages
    - Simple implementation
    - Easy training
    - Lower computation time
    - Easy to interpret
    - Overfitting can be reduced by regularization (lasso/ridge)
  + Disadvantages
    - May not be able to capture data properly (if correlation is not linear/uncorrelated)
      * Most linear regressions have low accuracy because relationship is not linear
    - Outliers have a huge effect on the linear regression’s performance and need to be treated beforehand
    - Linear regression assumes that no variables are correlated with each other
* Gradient descent
  + Iterative algorithm to minimize **cost function**
    - Cost function: measures difference between actual output and predicted output (error function)
    - Trains model to have optimized **weights**
      * Weights: numbers you use to turn samples into prediction (parameters used to predict output)



* + - Finding global minimum for cost function



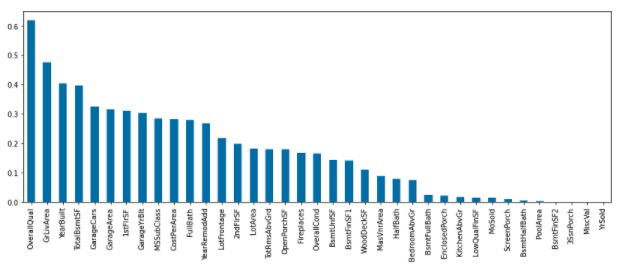
* + Alpha is learning rate, weights are w, derivative is gradient (searching for going down the mountain slope/learning rate)
  + So we used **stochastic gradient descent**
    - Uses single datapoint/example to calculate gradient and update weights
    - Shuffles dataset to get completely randomized
    - Advantages
      * Faster than other kinds of gradient descent
      * Less redundancy meaning less memory taken up (for larger databases this seems best)
    - Disadvantages
      * Very noisy, could lead algorithm away from minima and thus provide less accurate results (if it gets stuck in a local minima and not global)



* Decision trees
  + Predictive, based on binary rules
    - Decision making (sort of like yes-no/true-false path)
      * Asks these simple questions and depending on the answer, moves in a certain direction (probably like Akinator lol)
      * Separates these into different nodes, upon which more questions are asked to divide the data further until it cannot be divided anymore and thus a decision is made
  + Supervised learning model
    - Splitting is different for classification and regression trees
      * Accuracy is highly dependent on this
    - MSE usually used to decide whether to split a node in decision tree regression
      * Binary trees—picks value, splits data into two subsets, calculates MSE for each subset, chooses smallest MSE value
  + Advantages
    - Can be used for both classification and regression
    - Easy to interpret, understand, visualize
    - Output also easily understood
    - Less data preparation required, does not require normalization (works with categorical parameters), does not require scaling data
    - Not largely influenced by outliers/missing values
    - No assumptions about space distributions and classifier structure
  + Disadvantages
    - Overfitting but can be resolved by pruning/constraints
    - Cannot be used well with continuous numerical variables
    - Unstable (small changes in data => big change in structure)
    - Takes longer time to train model
    - Relatively expensive (time taken and complexity)
* Random forest
  + Recommendation system
  + Supervised learning algorithm used for regression and classification problems
  + Contains multiple decision trees and uses randomness to enhance accuracy
  + Advantages
    - Significantly more accurate than most nonlinear classifiers
    - Stable (sort of, takes average of all predictions and cancels out biases)
    - High flexibility
    - Easy to implement
    - Missing values aren’t an issue
  + Disadvantages
    - VERY slow, time-expensive
    - Harder to interpret than decision trees
* Ridge and lasso regression
  + Lasso regression
    - Lasso regression is linear regression but with restrictions
    - Similar to ridge but minimizes sum of squared residuals + (lambda \* sum of (coefficients squared))
      * Linear regression attempts to minimize only the sum of squared residuals
  + Ridge regression
    - Ridge regression is linear regression but with restrictions
    - Similar to lasso but minimizes sum of squared residuals + (lambda \* sum of (absolute value of coefficients))
  + When lambda = 0, the result is just linear regression
  + Both regressions introduce a little bias so that variance decreases, which in turn decreases the MSE
  + Advantages
    - Trades bias for variance and lowers MSE
  + Disadvantages
    - Becomes difficult to interpret coefficients in final model (get shrunk to zero)
  + Best to use either when interested in optimizing for predictions than inferences
  + Lasso regression is better when only a small number of predictor variables cause significant variance
  + Ridge regression is better when there are a lot of significant predictor variables (or variables of equal importance) – keeps all predictors in the model

|  |  |  |
| --- | --- | --- |
| **Type of Regression** | **Settings** | **Accuracy** |
| Linear | — | 0.8373695492356943 |
| Stochastic Gradient Descent | Max iterations – 30  Tol – 1e -3 | 0.8514499466140024 |
| Decision Trees | Max depth – 89 | 0.5675799643516808 |
| Random Forest | Number of trees – 60  Max depth – 18 | 0.7441594486181053 |
| Lasso Regression | Alpha – 1 | 0.8373824061492816 |
| Ridge Regression | Alpha – 30 | 0.8379786099550021 |

Stochastic Gradient Descent provided the highest accuracy, top down. However, when looking at the other kinds of regression, lasso and ridge regression also seem like convincing candidates to use for this dataset with optimized settings. Ridge regression provided the second highest accuracy, and according to my research, it is best to use it when there are several variables of equal importance. According to the mutual information feature selection…



There are several variables which have roughly equal importance when it comes to influencing house prices. So ridge regression works well. However, lasso regression was not far behind—at the end, one can see very insignificant values that could be removed (which lasso does but not ridge) so if alpha in lasso regression were increased, it is possible to get a higher accuracy.

Going back to SGD, however, this method may prove best overall with optimized settings. It works nicely for large databases like this one and doesn’t include the fallacy of assuming independent variables like linear regression, lasso regression, and ridge regression ~~at least I think so~~. Data preparation is required, unfortunately, but that versus the accuracy and amount of time taken results in a net benefit, unlike with Random Forest and Decision Trees (which take too much maneuvering with the settings and also are very time-consuming to train, especially with a large dataset).