**DSBA 6156**

**Assignment #2**

**Modules 4-7**

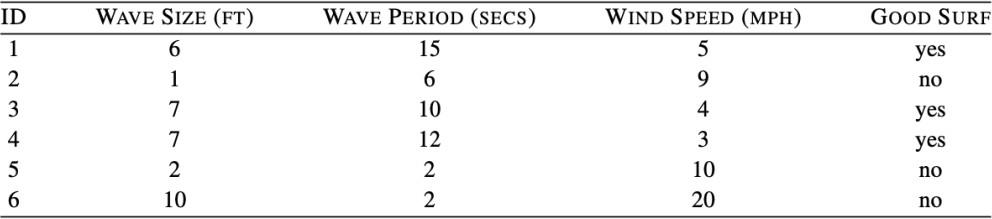
1. A data analyst building a k-nearest neighbor model for a continuous prediction problem is considering appropriate values to use for k.
   1. Instead of experimenting with k values, the analyst thinks it might be best to just set k to the total number of instances in the training set. Do you think that the analyst is likely to get good results using this value for k? Why / why not?

If you specified a number of k, you have a chance of losing sensitivity between variables as the locality is crucial when looking for specific nuances. Overgeneralization would also occur making any prediction be nearly the same rather to every new instance. It is also inefficient to use the whole dataset as it would be expensive to do it for every prediction.

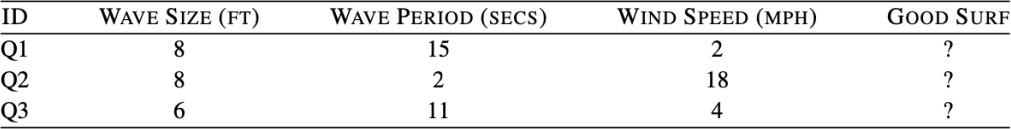
* 1. If the analyst was using a distance weighted KNN rather than a simple KNN for the predictions, would this have made the analyst’s idea any more useful? Why / why not?

Improved weighting causes the closer neighbors have a greater influence on the prediction than farther one creating a form of differentiation between instances based on their distance. However overgeneralization still persists the prediction would still be biased towards the central tendency of the target variable across the entire dataset, reducing the model's ability to capture local patterns thus still having inefficiency when calculating for all distances.

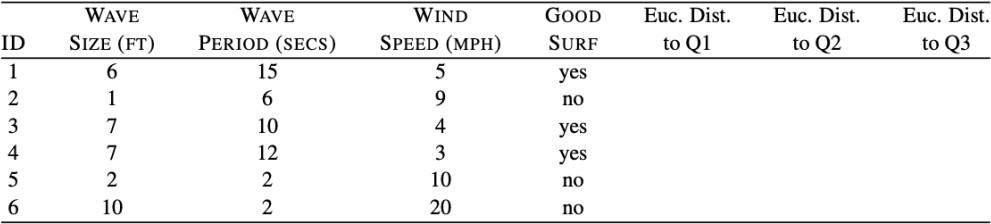
1. The table below lists a dataset that was used to create a nearest neighbor model that predicts whether it will be a good day to go surfing.

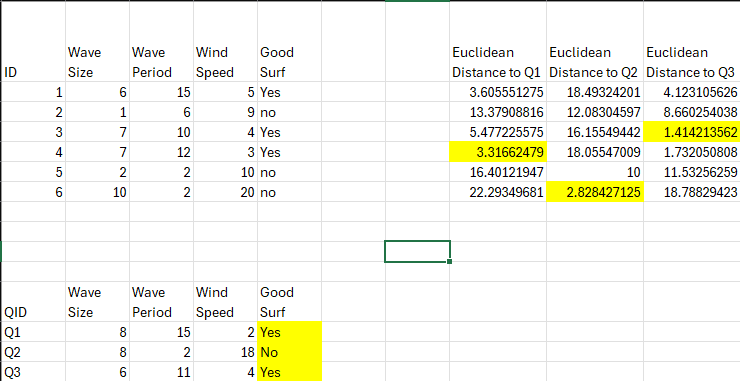


Assuming that the model uses Euclidean distance to find the nearest neighbor, what prediction will the model return for each of the following query instances?



Use Excel to limit manual calcs in the solution:





1. Linear Regression: A model is being built to predict the amount of oxygen that an astronaut consumes when performing five minutes of intense physical work. The descriptive features for the model will be the age of the astronaut and their average heart rate throughout the work. The regression model is:

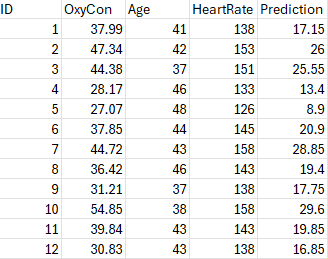
𝑶𝑿𝒀𝑪𝑶𝑵 = 𝒘[𝟎] + 𝒘[𝟏] × 𝑨𝑮𝑬 + 𝒘[𝟐] × 𝑯𝑬𝑨𝑹𝑻𝑹𝑨𝑻𝑬

The table that follows shows a historical dataset that has been collected for this task.

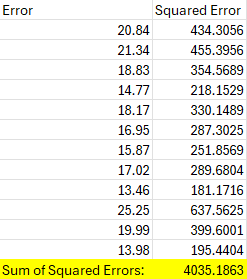
| ID | OXYCON | AGE | HEARTRATE |
| --- | --- | --- | --- |
| 1 | 37.99 | 41 | 138 |
| 2 | 47.34 | 42 | 153 |
| 3 | 44.38 | 37 | 151 |
| 4 | 28.17 | 46 | 133 |
| 5 | 27.07 | 48 | 126 |
| 6 | 37.85 | 44 | 145 |
| 7 | 44.72 | 43 | 158 |
| 8 | 36.42 | 46 | 143 |
| 9 | 31.21 | 37 | 138 |
| 10 | 54.85 | 38 | 158 |
| 11 | 39.84 | 43 | 143 |
| 12 | 30.83 | 43 | 138 |

Reference slides 40 - 47 in 7A lecture material, to solve (use Excel / other tool so you don’t have to do it by hand):

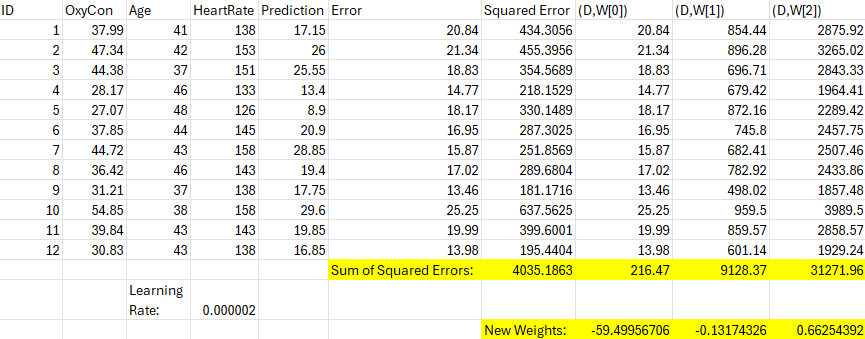
* 1. Assuming that the current weights in a multivariate linear regression model are w[0] = -59.50, w[1] = -0.15, and w[2] = 0.60, make a prediction for each training instance using this model.

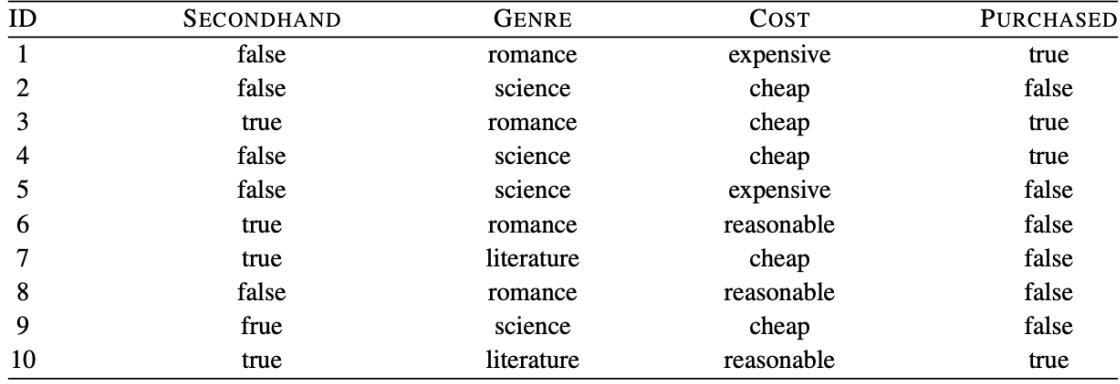


* 1. Extra Credit: Calculate the sum of squared errors for predictions generated in Part(a).



* 1. Extra Credit: Assuming a learning rate of 0.000002, calculate the weights at the next iteration of the gradient descent algorithm.



1. Naive-Bayes: The table below lists a dataset of books and whether or not they were purchased by an individual (i.e., the feature PURCHASED is the target feature in this domain).
   1. Assuming conditional independence between features given the target feature value, calculate the probability (rounded to four places of decimal) of each outcome (PURCHASED=true, and PURCHASED=false) for the following book:

SECONDHAND=*false,* GENRE=*literature,* COST=*expensive*

*P(True) = .5 \* .25 \* .25 \*.4 = .0125*

*P(False) = .5 \* .1667 \* .1667 \* .6 = .0083*

*a = .0125 + .0083 = .0208*

*P(True) = .0125/.0208 = .6010*

*P(False) = ..0083/.0208 = .3990*

* 1. What prediction would a naive Bayes classifier return for the above book?\

It returns outcome with the maximum probability with its prediction. For each instance the outcome PURCHASED=true is the MAP prediction and will be the outcome returned by a naive Bayesmodel

1. What happens to the bias (underfit) variance (overfit) tradeoff as we increase K in a KNN model? Is it more likely to underfit or overfit the training data?

The more we increase k, were likely to underfit as the model becomes more stable and less sensitive to the noise in the training data. When K is lsmaller, your more likely to overfit the making the model more complex.

1. Which model is more sensitive to feature scaling: KNN or Decision Tree? Why?

KNN is more sensitive to feature scaling as it uses different measures to determine the nearest neighbor, if one feature has a large range of values then it can oversaturate in the calculation making values weigh towards that one more than others

1. You are analyzing a KNN and a Decision Tree for runtime performance. Which would likely be faster once deployed for inference? Explain in terms of ‘eager’ vs ‘lazy’ learners.

Eager learners generalize from the training data before receiving data for prediction, meaning they are ready to make predictions instantly. Decision Trees are an example of eager learners.

Lazy Learners do not build the model in advance, they wait to receive data before they start anything to start generalizing. K-Nearest Neighbors (KNN) is an example of a lazy learner because it doesn't learn a discriminative function from the training data but memorizes the dataset instead.

Eager Learners are typivallyu faster because they already have the model created during trainfng then apply the model to make predictions.

1. Linear regression models can be adjusted to model non-linear relationships (between target and descriptive features) — True / False?

True

You can use polynomial terms, interaction terms, and logs to fit a much wider rance while the model itself stays linear, this is call polynomial regression.

1. Regardless of learning rate, gradient descent will always end up at the global minimum for linear regression models (learning rate just changes how quickly it will find the minimum) — True / False?

True

Gradient descent is guaranteed to converge to the global minimum, given enough time and assuming the learning rate is not too high.

If the learning rate is too large, the algorithm might overshoot the minimum and fail to converge, or it might diverge entirely. On the other hand, if the learning rate is too small, convergence will be very slow, which can be computationally inefficient. This is so the convex function can converge on a specific point.

learning rate needs to be chosen carefully to ensure convergence within a reasonable timeframe and to avoid divergence

1. Regularization techniques such as Ridge / Lasso Regression can be used to improve the performance of a model that is exhibiting poor training accuracy (<20%) — True / False?

Ridge and Lasso work to combat overfitting when the model you run has a high training accuracy but low test accuracy. If the model is showing low accuracy in the train (<20%), then this is a sign of underfitting making it not likely to improve the performance. Regularization is better when the model is too noisy/complex.

1. Outliers have little to no effect on the coefficients estimated by linear regression models

— True / False?

False, outliers can have a significant impact on the model coefficients as it can cause a large amount of skewing by falling far from the central data points. If there are outliers, the regression line will try to take account for it and pull the line towards them.