MACHINE LEARNING – WORKSHEET 2

1. In which of the following you can say that the model is overfitting?

Ans- A) High R-squared value for train-set and High R-squared value for test-set.

1. Which among the following is a disadvantage of decision trees?

Ans- B) Decision trees are highly prone to overfitting.

1. Which of the following is an ensemble technique?

Ans- C) Random Forest

1. Suppose you are building a classification model for detection of a fatal disease where detection of the disease is most important. In this case which of the following metrics you would focus on?

Ans- C) Precision

1. 5. The value of AUC (Area under Curve) value for ROC curve of model A is 0.70 and of model B is 0.85. Which of these two models is doing better job in classification?

Ans- D) Data Insufficient

1. Which of the following are the regularization technique in Linear Regression??

Ans- A) Ridge D) Lasso

1. Which of the following is not an example of boosting technique?

Ans- A) Adaboost D) Xgboost.

1. Which of the techniques are used for regularization of Decision Trees?

Ans- A) Pruning B) L2 regularization

1. Which of the following statements is true regarding the Adaboost technique?

Ans- A) We initialize the probabilities of the distribution as 1/n, where n is the number of data-points

B) A tree in the ensemble focuses more on the data points on which the previous tree was not performing well

1. Explain how does the adjusted R-squared penalize the presence of unnecessary predictors in the model?

Ans- Use adjusted R-squared to compare the goodness-of-fit for regression models that contain differing numbers of independent variables. Let’s say you are comparing a model with five independent variables to a model with one variable and the five variable model has a higher R-squared. Is the model with five variables actually a better model, or does it just have more variables? To determine this, just compare the adjusted R-squared values.

The adjusted R-squared adjusts for the number of terms in the model. Importantly, its value increases only when the new term improves the model fit more than expected by chance alone. The adjusted R-squared value actually decreases when the term doesn’t improve the model fit by a sufficient amount.

1. Differentiate between Ridge and Lasso Regression

Ans- Ridge and Lasso regression uses two different penalty functions. Ridge uses l2 where as lasso go with l1. In ridge regression, the penalty is the sum of the squares of the coefficients and for the Lasso, it’s the sum of the absolute values of the coefficients. It’s a shrinkage towards zero using an absolute value (l1 penalty) rather than a sum of squares(l2 penalty).

As we know that ridge regression can’t zero coefficients. Here, you either select all the coefficients or none of them whereas LASSO does both parameter shrinkage and variable selection automatically because it zero out the co-efficients of collinear variables. Here it helps to select the variable(s) out of given n variables while performing lasso regression.

Another type of regularization method is ElasticNet, it is hybrid of lasso and ridge regression both. It is trained with L1 and L2 prior as regularizer. A practical advantage of trading-off between Lasso and Ridge is that, it allows Elastic-Net to inherit some of Ridge’s stability under rotation.

1. What is VIF? What is the suitable value of a VIF for a feature to be included in a regression modelling?

Ans- A variance inflation factor(VIF) detects [multicollinearity](https://www.statisticshowto.com/multicollinearity/)in [regression analysis](https://www.statisticshowto.com/probability-and-statistics/regression-analysis/). Multicollinearity is when there’s [correlation](https://www.statisticshowto.com/probability-and-statistics/correlation-analysis/)between predictors (i.e. [independent variables](https://www.statisticshowto.com/independent-variable-definition/)) in a model; it’s presence can adversely affect your regression results. The VIF estimates how much the variance of a regression coefficient is inflated due to multicollinearity in the model.

The smallest possible value for VIF is 1, which indicates the complete absence of collinearity. Typically in practice there is a small amount of collinearity among the predictors. As a rule of thumb, a VIF value that exceeds 5 or 10 indicates a problematic amount of collinearity.

1. Why do we need to scale the data before feeding it to the train the model?

Ans- The input variables are those that the network takes on the input or visible layer in order to make a prediction.

A good rule of thumb is that input variables should be small values, probably in the range of 0-1 or standardized with a zero mean and a standard deviation of one.

Whether input variables require scaling depends on the specifics of your problem and of each variable.

You may have a sequence of quantities as inputs, such as prices or temperatures.

If the distribution of the quantity is normal, then it should be standardized, otherwise the data should be normalized. This applies if the range of quantity values is large (10s, 100s, etc.) or small (0.01, 0.0001).

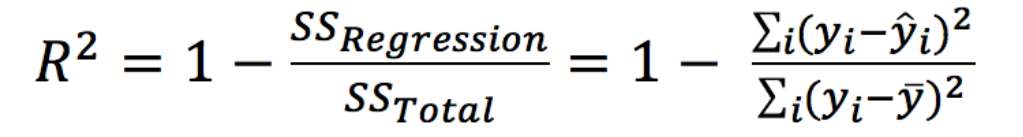
If the quantity values are small (near 0-1) and the distribution is limited (e.g. standard deviation near 1) then perhaps you can get away with no scaling of the data.

14. What are the different metrics which are used to check the goodness of fit in linear regression?

Ans- There are 3 main metrics for model evaluation in regression:

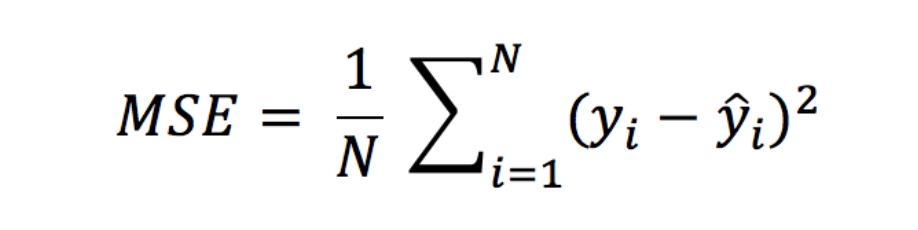
1. R Square/Adjusted R Square

R Square measures how much of variability in dependent variable can be explained by the model. It is square of Correlation Coefficient(R) and that is why it is called R Square.



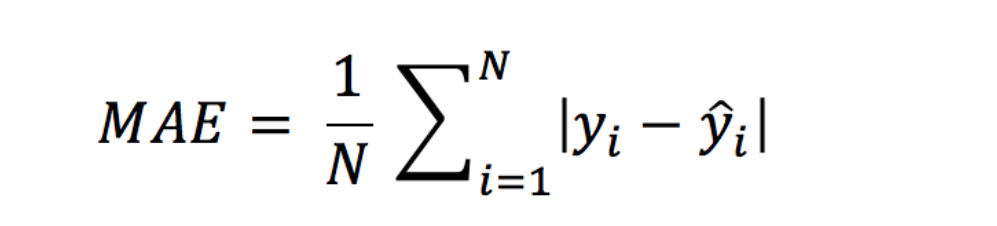
1. Mean Square Error(MSE)/Root Mean Square Error(RMSE)

While R Square is a relative measure of how well the model fits dependent variables, Mean Square Error is an absolute measure of the goodness for the fit.



1. Mean Absolute Error(MAE)

Mean Absolute Error(MAE) is similar to Mean Square Error(MSE). However, instead of the sum of square of error in MSE, MAE is taking the sum of absolute value of error.



 MSE gives larger penalisation to big prediction error by square it while MAE treats all errors the same.

15. From the following confusion matrix calculate sensitivity, specificity, precision, recall and accuracy.

|  |  |  |
| --- | --- | --- |
| Actual/Predicted | True | False |
| True | 1000 | 50 |
| False | 250 | 1200 |

Ans-

|  |  |
| --- | --- |
| **Measure** | **Calculated value** |
| Sensitivity (TP/TP+FN) | 1000/(1000+50)=0.95 |
| Specificity(TN/TN+FP) | 1200/ (1200+250)=0.82 |
| Precision(TP/TP+FP) | 1000/(1000+250)= 0.80 |
| Recall(TP/TP+FN) | 1000/(1000+50)=0.95 |
| Accuracy(TP+TN/TP+TN+FN+FP) | 1000+1200/1000+1200+250+50=0.88 |
|  |  |