**WORKSHEET**

**MACHINE LEARNING – WORKSHEET 2**

***In Q1 to Q5, only one option is correct, Choose the correct option:***

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

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

**B) Low R-squared value for train-set and High R-squared value for test-set.**

**C) High R-squared value for train-set and Low R-squared value for test-set.**

**D) None of the above**

**Answer :-**

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

**A) Decision trees are prone to outliers.**

**B) Decision trees are highly prone to overfitting.**

**C) Decision trees are not easy to interpret**

**D) None of the above.**

**Answer :-** B) Decision trees are highly prone to overfitting.

**3. Which of the following is an ensemble technique?**

**A) SVM B) Logistic Regression**

**C) Random Forest D) Decision tree**

**Answer :-** C) Random Forest

**4. 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?**

**A) Accuracy B) Sensitivity**

**C) Precision D) None of the above.**

**Answer :-** B) Sensitivity

**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?**

**A) Model A B) Model B**

**C) both are performing equal D) Data Insufficient**

**Answer :-** A) Model A

**In Q6 to Q9, more than one options are correct, Choose all the correct options:**

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

**A) Ridge B) R-squared**

**C) MSE D) Lasso**

**Answer :-** A) Ridge D) Lasso

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

**A) Adaboost B) Decision Tree**

**C) Random Forest D) Xgboost.**

**Answer :-** A) Adaboost D) Xgboost.

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

**A) Pruning B) L2 regularization**

**C) Restricting the max depth of the tree D) All of the above**

**Answer :- D) All of the above**

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

**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**

**C) It is example of bagging technique**

**D) None of the above**

**Answer :-**

**Q10 to Q15 are subjective answer type questions, Answer them briefly.**

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

**Answer :-** The adjusted R-squared is a modified version of R-squared that has been adjusted for the number of predictors in the model. The adjusted R-squared increases only if the new term improves the model more than would be expected by chance. It decreases when a predictor improves the model by less than expected by chance. The adjusted R-squared can be negative, but it’s usually not.  It is always lower than the R-squared.

**11. Differentiate between Ridge and Lasso Regression.**

**Answer :-**

**Ridge Regression :**  
Ridge regression is an extension for linear regression. It’s basically a regularized linear regression model. The λ parameter is a scalar that should be learned as well, using a method called **cross validation** that will be discussed in another post.

A super important fact we need to notice about ridge regression is that it enforces the β coefficients to be lower, but it **does not** enforce them to be zero. That is, it will not get rid of irrelevant features but rather**minimize their impact on the trained model.**

**Limitation of Ridge Regression:** Ridge regression decreases the complexity of a model but does not reduce the number of variables since it never leads to a coefficient been zero rather only minimizes it. Hence, this model is not good for feature reduction.

**Lasso Regression :**  
Lasso is another extension built on regularized linear regression, but with a small twist.The only difference from Ridge regression is that the regularization term is in absolute value. Lasso method overcomes the disadvantage of Ridge regression by not only punishing high values of the coefficients β but actually setting them to zero if they are not relevant. Therefore, you might end up with fewer features included in the model than you started with, which is a huge advantage.

Limitation of Lasso Regression:

* Lasso sometimes struggles with some types of data. If the number of predictors *(p)* is greater than the number of observations *(n)*, Lasso will pick at most n predictors as non-zero, even if all predictors are relevant (or may be used in the test set).
* If there are two or more highly collinear variables then LASSO regression select one of them randomly which is not good for the interpretation of data

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

**Answer :-** 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.

VIFs are usually calculated by software, as part of regression analysis. You’ll see a VIF column as part of the output.

What is known is that the more your VIF increases, the less reliable your regression results are going to be. In general, a VIF above 10 indicates high correlation and is cause for concern. Some authors suggest a more conservative level of 2.5 or above. Sometimes a high VIF is no cause for concern at all.

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

**Answer :-** Feature Scaling or Standardization: It is a step of Data Pre Processing which is applied to independent variables or features of data. It basically helps to normalise the data within a particular range. Sometimes, it also helps in speeding up the calculations in an algorithm.

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

**Answer :-** Three statistics are used in Ordinary Least Squares (OLS) regression to evaluate model fit: R-squared, the overall F-test, and the Root Mean Square Error (RMSE). All three are based on two sums of squares: Sum of Squares Total (SST) and Sum of Squares Error (SSE).

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

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

**Answer :- Sensitivity = (1000)/(1000+250) = 0.8**

**Specificity / Recall = (1200)/(50+1200) = 0.96**

**Precision = (1000)/(1000+50) = 0.952**

**Accuracy = (1000+1200)/(1000+50+250+1200) = 0.88**