/*
Author: Wang Jinge
Date: 3/12/2019
For 695-Machine Learning homework2.
*/

I.Please answer the following questions related to Machine Learning concepts:

1. [3 points] Explain what is the bias-variance trade-off?

In statistics and machine learning, the bias–variance tradeoff is the property of a set of predictive models whereby models with a lower bias in parameter estimation have a higher variance of the parameter estimates across samples, and vice versa. The bias–variance dilemma or problem is the conflict in trying to simultaneously minimize these two sources of error that prevent supervised learning algorithms from generalizing beyond their training set.

2. [3 points] Describe few techniques to reduce bias and variance respectively.

Reduce bias:

①Look for better features, more typical.

②Use more features, improve the complexity of model.

Reduce variance:

①Use more data.

②reduce the number of features.---reduce the complexity.

③Use regularization method.

④Use cross-validation.

3. [4 points] Describe following regression measures:

1) RMSE,

Root Mean Square Error (RMSE) is the standard deviation of the residuals (prediction errors). Residuals are a measure of how far from the regression line data points are; RMSE is a measure of how spread out these residuals are. In other words, it tells you how concentrated the data is around the line of best fit.

2) R2 measure

R-squared is a statistical measure of how close the data are to the fitted regression line. It is also known as the coefficient of determination, or the coefficient of multiple determination for multiple regression.

The definition of R-squared is fairly straight-forward; it is the percentage of the response variable variation that is explained by a linear model. Or:

R-squared = Explained variation / Total variation

R-squared is always between 0 and 100%:

0% indicates that the model explains none of the variability of the response data around its mean.

100% indicates that the model explains all the variability of the response data around its mean.

In general, the higher the R-squared, the better the model fits your data. However, there are important conditions for this guideline that I'll talk about both in this post and my next post.

4. [10 points] Explain following concepts related to classification measures:

1) confusion matrix,

A confusion matrix is a table that is often used to describe the performance of a classification model (or "classifier") on a set of test data for which the true values are known.

3) precision,

precision is the fraction of retrieved documents that are relevant to the query:

$$\mathbf{precision} = \frac{|\{\mathbf{relevant\ documents}\} \cap \{\mathbf{retrieved\ documents}\}|}{|\{\mathbf{retrieved\ documents}\}|}$$

For example, for a text search on a set of documents, precision is the number of correct results divided by the number of all returned results.

Precision takes all retrieved documents into account, but it can also be evaluated at a given cut-off rank, considering only the topmost results returned by the system. This measure is called precision at n or P@n.

Precision is used with recall, the percent of all relevant documents that is returned by the search. The two measures are sometimes used together in the F1 Score (or f-measure) to provide a single measurement for a system.

Note that the meaning and usage of "precision" in the field of information retrieval differs from the definition of accuracy and precision within other branches of science and technology.

4) recall,

$$\mathrm{recall} = \frac{|\{\mathrm{relevant\ documents}\} \cap \{\mathrm{retrieved\ documents}\}|}{|\{\mathrm{relevant\ documents}\}|}$$

For example, for a text search on a set of documents, recall is the number of correct results divided by the number of results that should have been returned.

In binary classification, recall is called sensitivity. It can be viewed as the probability that a relevant document is retrieved by the query.

It is trivial to achieve recall of 100% by returning all documents in response to any query. Therefore, recall alone is not enough but one needs to measure the number of non-relevant documents also, for example by also computing the precision.

5) F1 score,

F1 Score is the weighted average of Precision and Recall. Therefore, this score takes both false positives and false negatives into account. Intuitively it is not as easy to understand as accuracy, but F1 is usually more useful than accuracy, especially if you have an uneven class distribution. Accuracy works best if false positives and false negatives have similar cost. If the cost of false positives and false negatives are very different, it’s better to look at both Precision and Recall.

6) ROC curve.

The ROC curve is a fundamental tool for diagnostic test evaluation.

In a ROC curve the true positive rate (Sensitivity) is plotted in function of the false positive rate (100-Specificity) for different cut-off points of a parameter. Each point on the ROC curve represents a sensitivity/specificity pair corresponding to a particular decision threshold. The area

under the ROC curve (AUC) is a measure of how well a parameter can distinguish between two diagnostic groups (diseased/normal).

II. Please answer the following questions related to Concept Learning:

5. [10 points] Consider the following training examples (which is similar to EnjoySpt but with slightly different attributes) and the hypothesis space H that we described in lecture 3-2 (i.e., hypothesis is conjunction of attributes).

	Temp	Humidity	Water	Sky	EnjoySport
1	Warm	Normal	Warm	Sunny	Yes
2	Warm	Normal	Cold	Sunny	Yes
3	Cold	Normal	Warm	Rainy	No
4	Warm	High	Warm	Sunny	Yes

Trace the Candidate-Elimination algorithm to show the sequence of S and G boundary set