MACHINE LEARNING
BARUCH COLLEGE
SPRING 2015
MIGUEL A. CASTRO
T.A.: LIQUN ZHU

Homework Assignment 3

1. As in class, suppose you are confronted with a security's return series Y and you build a neural network model to predict the series. You've used regularization techniques and are reasonably assured that your model generalizes well, and that, therefore, its R^2 is very close to the optimum, $maxR^2$. We saw in class that if you use the model's return prediction to build a trading strategy, and assume that the expected return of your strategy is proportional to the volatility of the predictable portion of the return series, $r = \lambda\sigma_Y$, then the strategy's Sharpe Ratio is:

$$SR \simeq \lambda R.$$

Confirm that this is true by using the fact that the volatility in the denominator of the Sharpe Ratio is the full (noisy) signal's volatility and not just the volatility of the predictable (target) signal. Note also that this is the *unadjusted* Sharpe Ratio (*i.e.,* we are not subtracting the risk-free rate from the expected return).

2. a. Show that the Random Classifier traces the Random Classifier Line in the ROC curve.

   b. Show that the Youden Index, *J*, of a given classifier is equivalent to the Euclidean Distance from the given classifier to the Random Classifier Line. In other words, *J* always ranks any two classifiers in exactly the same way as the Euclidean Distance to the Random Classifier Line would.

   c. Show that the Euclidean Distance Accuracy, *DistAcc*, the normalized Euclidean distance from a given classifier to the Perfect Classifier, is not equivalent to the distance from the given classifier to the Random Classifier Line.

   d. Construct two example classifiers A and B (as coordinates on a ROC plot) such that the Euclidean Distance Accuracy, *DistAcc* will rank A as better than B, while the Cost-Adjusted Euclidean Distance Accuracy, with a given *W,* will revert the ranking found using *DistAcc*. Explain how this can happen.

3. In class we considered a special classifier used to decide whether to enter a trade. Class 1 represents "The Trade is Profitable" and Class 0 represents "The Trade is Unprofitable." The profit of entering a

profitable trade is P while the loss of entering an unprofitable trade is L. The classifier's output o ranges from 0 to 1 and can be considered a proxy for the probability that Class 1 is true. From this, we saw that setting the classifier threshold θ to

$$\theta = \frac{L}{L+P}$$

was equivalent to entering the trade when the expected profit exceeded the expected loss. We also noted that in HFT, L is usually larger than P. What simple condition does this impose on θ?

4. Using the same classifier as in 2. above, explain the following two equations:
   - $Expected\ Profit = TP \cdot P$;
   - $Expected\ Loss = FP \cdot L + FN \cdot P$.

   Here we have assumed that TP is the fraction of all classifications that were correctly identified as Class 1 by the classifier, FP is the fraction of al classifications that were incorrectly identified as Class 1, and FN is the fraction of all classifications that were incorrectly identified as Class 0. Again requiring that the Expected Profit exceed the Expected Loss, we get a relation between classifier performance (the left-hand side, in terms of TP, FN, FP) and the Profit/Loss ratio (in the right-hand side):

   $$\frac{(TP - FN)}{FP} > \frac{L}{P}.$$

   Discuss why this relation makes sense. (What happens as the Loss becomes larger in relation to the Profit? What happens to the classifier performance as TP increases, etc?)

   Discuss how this relation compares with the Youden Index as a classifier performance measure.

   Discuss why, in this particular case, the True Negatives (TN) do not enter the expression for classifier performance (the right-hand side of the above equation).

   Given that $L > P$, are False Positives (Type I Errors) more or less costly than False Negatives (Type II Errors)? Explain.

5. You are designing a classifier to predict whether a trade will yield +2, +1, 0, -1, or -2 cents per stock traded. This type of classifier can be called an "*Ordinal Classifier*" because the classes can be ranked in order, where class +2 (representing a trade that yields a profit of 2 cents) is more distant from class -1 (representing a trade that yields a loss of 1 cent), than it is to class +1. You trained two classifiers, **A** and **B**, and tested each of them on the same hold-out data set, which produced the following two (count) Confusion Matrices

		Actual				
		-2	-1	0	+1	+2
Predicted	-2	1237	251	183	127	76
	-1	213	958	236	178	82
	0	172	212	1102	280	101
	+1	103	188	247	1007	275
	+2	65	79	124	289	984

(Confusion Counts for Classifier **A**)

		Actual				
		-2	-1	0	+1	+2
Predicted	-2	1206	150	84	26	375
	-1	113	989	186	78	281
	0	272	112	1051	230	151
	+1	203	138	197	1055	255
	+2	365	29	24	39	987

(Confusion Counts for Classifier **B**)

Ignoring transaction costs, discuss how you would decide which classifier is better. Would a diagonality measure be useful as the sole criterion for comparison? Why or why not? How many pair-wise binary classifiers would you have to analyze if you were to make a pair-wise comparison among the classes? Would you treat all such pair-wise comparisons the same way to arrive at your overall comparison between classifiers **A** and **B**? How many pair-wise comparisons would you have to make if you had *N* classes instead of 5? Would your analysis change if you added asymmetric transaction costs (where entering a trade is always more expensive than not trading)? Why or why not?

6. Discuss the advantages and disadvantages of the five Data Imputation methods that we discussed within our European Companies Case Study.

7. Suppose you're searching for Association Rules in a transaction database with 1,000,000 transactions. Your algorithm is using the Support-Confidence Framework, and you have decided to require a Support lower bound of $s = 1\%$ and a confidence level of no less than $c = 10\%$. You're considering Itemsets X and Y and have measured the following occurrences in the database: Number of occurrences of X is 500,000; Y occurs 400,000 times; X and Y *together* occur 200,000 times. (a) What is the support of the rule $X \Rightarrow Y$? Does it meet the support threshold? (b) What is the confidence of the rule $X \Rightarrow Y$? Does it meet the confidence threshold? (c) Based on your answers to (a) and (b), is this an Interesting Association Rule according to the Support-Confidence Framework? Discuss whether this is a good idea. Now suppose that X and Y occur together only 2,000 times. (d) What is the support of the rule $X \Rightarrow Y$? Does it meet the support threshold? (e) What is the confidence of the rule $X \Rightarrow Y$ and does it meet the confidence threshold? (f) Based on the answers to (e) and (f), is this an Interesting Association Rule according to the Support-Confidence Framework? Discuss whether this is a good idea.