Topics in Quantitative Finance (FIN 528)
Machine Learning for Finance
Midterm Exam

Instructor: Jaehyuk Choi

2017-18 Module 3 (2018. 4. 9)

1. (4 points) Answer each of the following TRUE or FALSE.
   (a) The training accuracy of KNN with $K = 1$ is always 100%.
   (b) The batch gradient descent is always better than stochastic gradient because it makes use of all training samples.
   (c) The linear regression can be performed with either matrix algebra or using gradient descent.
   (d) Regularization reduces both bias and variance.

Solution:
(a) True
(b) False
(c) True
(d) False

2. (3 points) **(Maximum Likelihood, Bayesian)** If a coin is fair, the probability for head (vs tail) is 50% (vs 50%). If a coin is faulty, it is double-headed, so the probability for head (vs tail) is 100% (vs 0%).
   (a) If your coin is fair, what is the probability of getting 10 heads out of 10 tosses?
   (b) You withdrew one coin from People's Bank of China (PBoC), tossed it 10 times and got 10 heads. You cannot look at both sides of coin, so you cannot tell whether your coin is faulty or not. Assuming that PBoC has probability p of minting faulty coin, what is the likelihood of your observation (i.e., 10 heads)? What is the value of p which maximize the likelihood?
   (c) This time assume you know that $p = 1/1000$. Under the same observation (i.e. 10 heads from a randomly selected coin), what is the probability that your coin is faulty?

Solution:
(a) $1/2^{10} = 1/1024 \approx 0.1\%$.

(b) The likelihood is given as

$$l(p) = p \cdot 1 + (1-p) \cdot \frac{1}{1024} = \frac{1 + 1023p}{1024},$$

which is maximized when $p = 1$.

(c)

$$\text{Prob} = \frac{p \cdot 1}{p \cdot 1 + (1-p)/1024} = \frac{1}{1 + 999/1024} \approx 50\%$$

3. (3 points) **(From ISLR Exercise 2.2)** Explain whether each scenario is a classification or regression problem, and indicate whether we are most interested in inference or prediction. Finally, provide n (sample size or number of observations) and p (number of features). (**Inference** is the process of understanding the properties of the underlying distribution or the relationship between input and output.)
   (a) We collect a set of data on the top 300 firms in China. For each firm we record profit, number of employees, industry and the CEO salary. We are interested in understanding which factors affect CEO salary.
   (b) We are considering launching a new product and wish to know whether it will be a success or a failure. We collect data on 20 similar products that were previously launched. For each product we have recorded whether it was a *success* or *failure*, price charged for the product, marketing budget, competition price, and ten other variables.
   (c) We are interest in predicting the % change in the USD/CNY exchange rate (e.g., USD/CNY = 6.3) in relation to the weekly changes in the stock markets in US and China. Hence we collect weekly data for all of 2017. For each week we record the % change in the USD/CNY, the % change in the US market, the % change in the Chinese market.

Solution:

(a) Regression. Inference. $n = 300$, $p = 3$.

(b) Classification. Prediction. $n = 20$, $p = 13$.

(c) Regression. Prediction. $n = 52$, $p = 2$.

4. (3 points) **(From ISLR Exercise 2.7)** The table below provides a training data set containing six observations, three predictors, and one qualitative response variable.

Obs.	X_1	X_2	X_3	Y
1	0	3	0	Red
2	2	0	0	Red
3	0	1	3	Red
4	0	1	2	Green
5	-1	0	1	Green
6	1	1	1	Red

Suppose that we wish to use this data set to make a prediction for Y when $X_1 = X_2 = X_3 = 0$ using K-nearest neighbors.

(a) Compute the Euclidean distance between each observation and the test point, $X_1 = X_2 = X_3 = 0$.

(b) What is our prediction with $K = 1$? Why?

(c) What is our prediction with $K = 3$? Why?

Solution:

(a)

Obs.	Distance	Y
1	3	Red
2	2	Red
3	$\sqrt{10}$	Red
4	$\sqrt{5}$	Green
5	$\sqrt{2}$	Green
6	$\sqrt{3}$	Red

(b) **Green** because the nearest neighbor is Obs. 5 (Green).

(c) **Red** because the 3 nearest neighbor are Obs. 2 (Red), 5 (Green) and 6 (Red)

5. (3 points) **(Gradient descent update rule)** Remind that, in linear regression and other models, the update rule for the coefficients in gradient descent is given as

$$\Delta w_j = \eta \sum_i (y_i - \hat{y}_i) X_{ij},$$

where η is learning rate, X_{ij} is the input from i-th sample for j-th feature, y_i is the i-th true response, and $\hat{y}_i$ is i-th prediction.

Imagine that you would like to use a quadratic polynomial regression for the input with two features:

$$Y \sim w_0 + w_{11} X_1 + w_{12} X_2 + w_{21} X_1^2 + w_{22} X_2^2 + w_{23} X_1 X_2$$

Write the update rule for the six coefficients, w_0, w_{11}, w_{12}, w_{21}, w_{22}, and w_{23}.

Solution: You can think of the polynomial regression as a linear regression with extended six features, $(1, X_1, X_2, X_1^2, X_2^2, X_1 X_2)$ and simply use the formula above:

$$\Delta w_0 = \eta \sum_i (y_i - \hat{y}_i)$$

$$\Delta w_{1j} = \eta \sum_i (y_i - \hat{y}_i) X_{1j} \quad \text{for} \quad j = 1,\ 2$$

$$\Delta w_{2j} = \eta \sum_i (y_i - \hat{y}_i) X_{2j}^2 \quad \text{for} \quad j = 1,\ 2$$

$$\Delta w_{23} = \eta \sum_i (y_i - \hat{y}_i) X_1\, X_2$$

6. (4 points) **(Evaluation metric, confusion matrix)** Didi Chuxing has been developing an automatic driving algorithm. One of the most important parts of the algorithm is to recognize humans in the image from camera to avoid accident. Imagine that they obtained the following two confusion matrices for different classification thresholds (low and high) from the algorithm.

Low Threshold Human?		Predicted P^*	Predicted N^*	Total
Actual	P	18	2	20
	N	50	30	80
	Total	68	32	100

High Threshold Human?		Predicted P^*	Predicted N^*	Total
Actual	P	10	10	20
	N	20	60	80
	Total	30	70	100

(a) Compute accuracy, recall (= true positive rate), precision, and false positive rate from each matrix.

(b) Which evaluation metric should be used in order to minimize accidents? Why?

(c) Compute the AUC score. (You can get only two points on the ROC curve from the matrices above. Use linear interpolation for the rest.)

Solution:

(a)

	Accuracy	Recall (TPR)	Precision	FPR
Low Threshold	48%	90%	26.5% (=18/68)	62.5% (=50/80)
High Threshold	70%	50%	33.3% (=10/30)	25% (=20/80)

(b) To avoid accidents, we need to minimize false negative, FN (i.e., actual human but NOT classified as human). Therefore recall (=TPR) should used as a metric. The low threshold is better (in recall) although overall accuracy is worse.

(c) The two ROC curves shown below, which pass the points, (1/4,1/2) and (5/8, 9/10), are accepted as solutions. The AUCs are 68.125% (=109/160) and 72.448% (=1391/1920) respectively.