CSE 575: Statistical Machine Learning: Mid-Term 1

Instructor: Prof. Hanghang Tong

February 16th, 2017

First Name:			
Last Name:			
Email:			
ASU ID:			
Q	Topic	Max Score	Score
1	MLE	15	
2	Bayes Classifer	25	
3	Naive Bayes	25	
4	Decision Boundary of 1NN	15	
5	Distance Metric and 1NN	20	
Total:		100	

- This exam book has **10** pages, including this cover page.
- You have 75 minutes in total.
- Good luck!

1 Maximum Likelihood Estimation (15 points)

[4 points.] Suppose we flip a coin, and observe either a head or a tail. The probability of observing a head in each trial is p $(0 \leq p \leq 1)$. If we flip the coins twice, and observe *(head, head)*, what is the maximum likelihood estimation of p? Justify your answer.

Solution: likelihood is p^2, and $p = 1$ maximizes the likelihood.

[4 points.] Suppose we flip a coin, and observe either a head or a tail. The probability of observing a head in each trial is p $(0 \leq p \leq 1)$. If we flip the coins twice, and observe *(head, tail)*, what is the maximum likelihood estimation of p? Justify your answer.

Solution: likelihood is $p(1-p)$. take the logorithm, and calculate its derivative, and set it as zero, we have

$$1/p - 1/(1-p) = 0$$

which gives $p = 0.5$.

[4 points.] Suppose we flip a coin, and observe either a head or a tail. The probability of observing a head in the i^{th} trial is $i * p$ $(i = 1, 2)$. If we flip the coins twice, and observe *(head, head)*, what is the maximum likelihood estimation of p? Justify your answer.

Solution: likelihood is $p * 2p$. In the meanwhile, we have $0 \leq 2p \leq 1$. $p = 0.5$ gives the maximum likelihood

[3 points.] Suppose we flip a coin, and observe either a head or a tail. The probability of observing a head in the i^{th} trial is $i * p$ $(i = 1, 2)$. If we flip the coins twice, and observe *(head, tail)*, what is the maximum likelihood estimation of p? Justify your answer.

Solution: likelihood is $p * (1 - 2p)$. take the logorithm, and calculate its derivative, and set it as zero, we have $p = 0.25$.

2 Bayes Classifier (25 points)

Continuous Bayes Classifier. We want to build a Bayes Classifier for a binary classification task ($y = 1$ or $y = 2$) with a 1-dimensional input feature (x). We know the following quantities: (1) $P(y = 1) = 0.4$; (2) $P(x|y = 1) = 0.5$ for $0 \leq x \leq 2$ and $P(x|y = 1) = 0$ otherwise; and (3) $P(x|y = 2) = 0.25$ for $0 \leq x \leq 4$ and $P(x|y = 2) = 0$ otherwise.

- **[3pts]**. What is the prior of the class label $y = 2$?

  $P(y = 2) = 0.6$

- **[2pts]**. What is $P(y = 1|x)$?

  $P(y = 1|x) = 4/7 \;\; 0 \leq x \leq 2$ and 0 otherwise $P(y = 2|x) = 3/7 \;\; 0 \leq x \leq 2 \; P(y = 2|x) = 1 \;\; 2 \leq x \leq 4$ and 0 otherwise

- **[2pts]**. For $x = 1$, what is class label your classifier will assign? What is the risk of this decision?

  $y = 1$, risk is $3/7$ (1 pt for each)

- **[2pts]**. For $x = 3$, what is class label your classifier will assign? What is the risk of this decision?

  $y = 2$, risk is 0 (1 pt for each)

- **[2pts]**. What is the decision boundary of your Bayes classifier?

  $[0, 2] : \; y = 1$; $(2, 4] : \; y = 2$; tie/unknown otherwise. (1 pt for each segment. lose 0.5pt if mistaken on the boundary)

- **[2pts]**. What is the Bayes error of your Bayes classifier?

  0.3 (0.15 x 2)

Discrete Bayes Classifier. We want to build a Bayes Classifier for a binary classification task ($y = 1$ or $y = 2$) with two binary features (x_1 and x_2). We know the following quantities: (1) $P(y = 1) = 0.2$; (2) $P(x_1 = 0, x_2 = 0|y = 1) = 0.3$, $P(x_1 = 0, x_2 = 1|y = 1) = 0.4$, $P(x_1 = 1, x_2 = 0|y = 1) = 0$, and $P(x_1 = 1, x_2 = 1|y = 1) = 0.3$; and (3) $P(x_1 = 0, x_2 = 0|y = 2) = 0.2$, $P(x_1 = 0, x_2 = 1|y = 2) = 0.3$, $P(x_1 = 1, x_2 = 0|y = 2) = 0.3$, and $P(x_1 = 1, x_2 = 1|y = 2) = 0.2$;

- **[2pts]**. What is the prior of the class label $y = 2$?

  $P(y = 2) = 0.8$

- **[2pts]**. What is $P(y = 2|x_1, x_2)$?

  $P(y = 2|x_1 = 0, x_2 = 0) = 8/11$, $P(y = 2|x_1 = 0, x_2 = 1) = 3/4$, $P(y = 2|x_1 = 1, x_2 = 0) = 1$, $P(y = 2|x_1 = 1, x_2 = 1) = 8/11$ (0.5pt for each case).

- **[2pts]**. For an example with the following features $x_1 = 0, x_2 = 1$, what is class label your classifier will assign? What is the risk of this decision?

  $y = 2$, risk is 0.25 (1 pt for each)

- **[2pts]**. For an example with the following features $x_1 = 1, x_2 = 0$, what is class label your classifier will assign? What is the risk of this decision?

  $y = 2$, risk is 0 (1 pt for each)

- **[2pts]**. What is the decision boundary of your Bayes classifier?

  all for $y = 2$ (0.5 pt for each case). alternatively, you can say 'no decision boundary' (since all the space is for $y = 2$).

- **[2pts]**. What is the Bayes error of your Bayes classifier?

  0.2

3 Naive Bayes Classifier (25 points)

Given the training data set in Figure 1, we want to train a binary classifier. In the figure, (1) the last column is the binary class label; (2) each of the first four columns is a binary feature, and (3) each row is a training example.

Input Feature $X = (x_1, x_2, x_3, x_4)$				Class Label Y
x_1	x_2	x_3	x_4	Y
1	1	0	0	1
0	1	1	0	1
1	0	1	1	1
1	1	0	0	1
1	0	0	1	0

Figure 1: Training Data Set

[5 points.] If we want to train a **Bayes Classifier**, how many *independent* parameters are there in your classifier? Justifiy your answer.

Solutions: $P(y = 1)$, for each class label, we have $2^4 - 1$ independent parameters. Total $2 * (2^4 - 1) + 1 = 31$ independent parameters.

[5 points.] If we want to train a **Naive Bayes Classifier**, how many *independent* parameters are there in your classifier? Justifiy your answer.

Solutions: (1) $P(y = 1)$, (2) for each class lable and each dimension of the feature, one independent parameter $P(x_i = 1|y = i)$ $(i = 1, 0)$. Total $2 * 4 + 1 = 9$ independent parameters.

[4 points.] Using the standard MLE (maximum likelihood estimation) to train a **Naive Bayes Classifier**, what is your estimation for $P(Y = 0)$?

Solutions: $P(Y = 0) = 0.2$.

[3 points.] Using the standard MLE (maximum likelihood estimation) to train a **Naive Bayes Classifier**, what is your estimation for $P(x_3 = 1|Y = 0)$?

Solutions: $P(x_3 = 1|Y = 0) = 0$.

[3 points.] Using the standard MLE (maximum likelihood estimation) to train a **Naive Bayes Classifier**, what is your estimation for $P(x_2 = 1|Y = 1)$?

Solutions: $P(x_2 = 0|Y = 1) = 0.75$.

[5 points.] Suppose we use the standard MLE (maximum likelihood estimation) to train a **Naive Bayes Classifier**. Now given a test example $X = (1, 0, 1, 1)$, what is the class label your classifier will predict? Justify your answer.

Solutions: $P(Y = 0|X) \propto P(Y = 0)P(x_1 = 1|Y = 0)P(x_2 = 0|Y = 0)P(x_3 = 1|Y = 0)P(x_4 = 1|Y = 0) = 0$ (due to $P(x_3 = 1|Y = 0) = 0$the above answer). In the meanwhile, $P(Y = 1|X) \neq 0$. Therefore, the classifier will predict as the positive example (i.e., $Y = 1$).

4 The Decision Boundary for 1NN (i.e., *1*-Nearest Neighbors Classifier) (15 points)

For each of the following figures, we are given a few data points in the 2-d space, each of which is labeled as either '+' or '-'. Draw the decision boundary for 1NN, assuming we use L_2 distance. (**5 points for each case**).

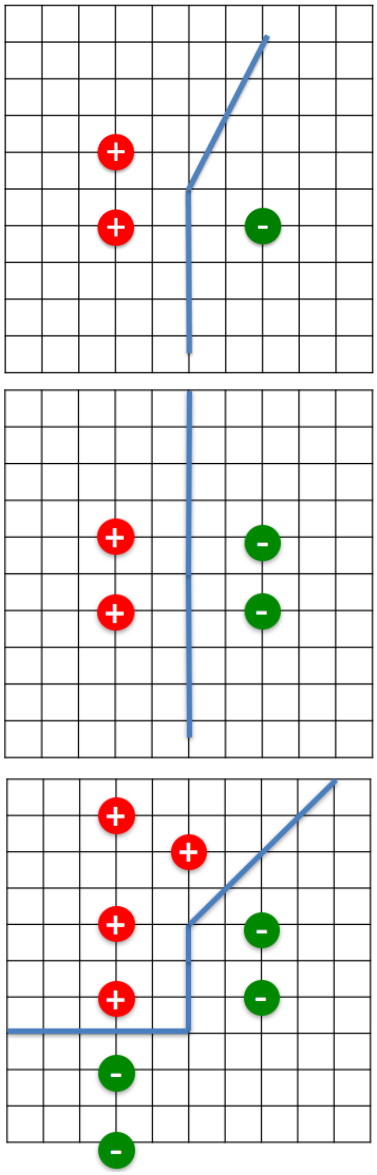

Figure 2: Training Data Set for 1NN Classifier

5 1NN Classifier and Distance Metric (20 points)

[5 points]. Given two training data points: (1) a positive example at $(0, 1)$ and (2) a negative example at $(0, -1)$, what is the decision boundary of 1NN classifier if we use $\mathbf{L_0}$ **distance**? Justify your answer.

Solution: everything except two horizonal lines: $y = 1$ and $y = -1$. If your answer is these two lines, get 2pts.

[5 points]. Given two training data points: (1) a positive example at $(0, 1)$ and (2) a negative example at $(0, -1)$, what is the decision boundary of 1NN classifier if we use $\mathbf{L_1}$ **distance**? Justify your answer.

Solution: the x-axis ($y = 0$).

[5 points]. Given two training data points: (1) a positive example at $(0, 1)$ and (2) a negative example at $(0, -1)$, what is the decision boundary of 1NN classifier if we use $\mathbf{L_2}$ **distance**? Justify your answer.

Solution: the x-axis ($y = 0$).

[5 points]. Given two training data points: (1) a positive example at $(0, 1)$ and (2) a negative example at $(0, -1)$, what is the decision boundary of 1NN classifier if we use $\mathbf{L_\infty}$ **distance**? Justify your answer.

Solutions: shown in the following figure. You also need to specify the (1) the functions for each of the four lines ($y = x + 1$,$y = x - 1$,$y = -x + 1$,$y = -x - 1$) and (2) two endpoints $(1, 0)$ and $(-1, 0)$. 0.5pt for each missed line or endpoint.

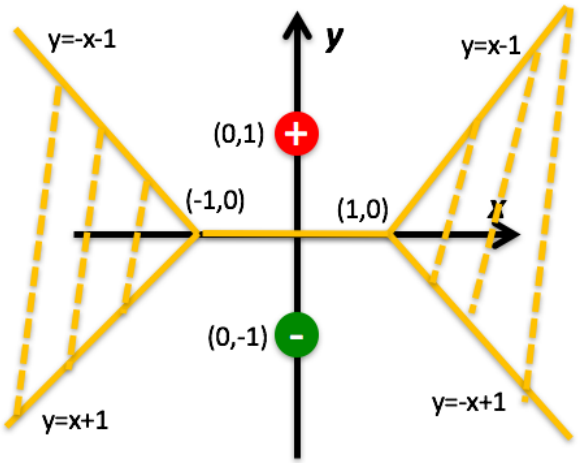

Figure 3: Decision Boundary of 1NN Classifiers with L_∞ Distance.