Pattern Classification and Recognition
ECE 681

Spring 2019
Homework #3: Curse of Dimensionality
(In-Class Assignment)

Due: 5:00 PM, Thursday, February 14, 2019
Grace Period Concludes: 11:30 PM, Tuesday, February 19, 2019

This homework assignment is worth **170 points**.
Each problem is worth some multiple of 10 points, and will be scored on the below letter scale.
The letter grades B through D may be modified by + (+3%) and A through D may be modified by a - (-3%).

- A+ = 100%: Exceeds expectations, and no issues identified
- A = 95%: Meets expectations, and (perhaps) minor/subtle issues
- B = 85%: Issues that need to be addressed
- C = 75%: Significant issues that must be addressed
- D = 65%: Major issues, but with noticeable perceived effort
- F = 50%: Major issues, and insufficient perceived effort
- Z = 30%: Minimal perceived effort
- N = 0%: Missing, or no (or virtually no) perceived effort

Your homework is not considered submitted until both components (**one self-contained pdf file** and your code) have been submitted. Please do not include a print-out of your code in the pdf file.

Distinguishing Pennies from Quarters

Consider the following (impossible) machine learning problem: We want to determine whether any particular coin is a quarter or a penny. The coin is in another room, where someone else will flip it some number of times and tell us the result. That is, the observations we get in order to classify the coin as either a penny or a quarter are a series of coin flip results, heads or tails.[1]

To learn a model (train a classifier) for this problem, we get P pennies and Q quarters ($N = P + Q$). We flip each one D times. The resulting dataset might look like this:

$$X = \begin{bmatrix} \mathrm{T} & \mathrm{H} & \dots & \mathrm{T} \\ \mathrm{H} & \mathrm{T} & \dots & \mathrm{T} \\ \vdots & \vdots & \vdots & \vdots \\ \mathrm{H} & \mathrm{H} & \dots & \mathrm{H} \end{bmatrix} \qquad Y = \begin{bmatrix} \mathrm{q} \\ \mathrm{q} \\ \vdots \\ \mathrm{p} \end{bmatrix}$$

where X contains the results of the coin flips and is of size $N \times D$ (coins $\times$ flips), and Y contains the truth as to which coin was flipped and is of size $N \times 1$ (coins $\times$ 1). The "data" (X) are the coin flip results: heads (H) or tails (T). The "truth" or "labels" (Y) are either quarter (q) or penny (p).

We can encode these data and associated labels by assigning 0 to represent tails and 1 to represent heads, and 0 to indicate the coin flipped was a penny and 1 to indicate the coin flipped was a quarter. With this encoding, the data and labels are:

$$X = \begin{bmatrix} 0 & 1 & \dots & 0 \\ 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & \dots & 1 \end{bmatrix} \qquad Y = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 0 \end{bmatrix}$$

[1] We will assume all the coins are perfectly fair, that is there is no dynamical bias in these coin flips (*i.e.*, $p(\mathrm{H}) = 0.5$ and $p(\mathrm{T}) = 0.5$). See "Dynamical Bias in the Coin Toss" by Persi Diaconis, Susan Holmes, and Richard Montgomery for a discussion of dynamical bias in coin flips (posted to Sakai under Syllabus and Additional Resources → Interesting Articles).

(40) 1. Consider first the case where each coin is flipped once ($D = 1$).

(a) Consider the special case $P = 1$ and $Q = 1$, where we flip one of each coin exactly once. Then the question "are these data separable?" is equivalent to the question "did the two coins produce different results?" because clearly we cannot begin try to tell which coin is which if the only features they show are identical. If they produce different results, however, we could learn, for example, that quarters usually land heads, while pennies usually land tails.
What is the probability that the two flips produce different results?

Solution: 1/2

For each penny and each quarter, each flip could produce 2 possible results, thus we have $2 \times 2 = 4$ possible results.

$\begin{bmatrix} P \\ Q \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$

✓ ✓

when $\begin{bmatrix} P \\ Q \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, they are different. Therefore the answer is $\frac{2}{4} = \frac{1}{2} = 0.5$

(b) Now consider the case where $P = 1$ and $Q = 2$. Note that the events 1) *in which the first quarter and the penny produce the same result*, and 2) *in which the second quarter and the penny produce the same result* are independent. What is the probability that neither quarter produces the same result as the penny? $1/4 = 0.25$

Solution: for each coin, each flip have 2 possible results.
Total possible results should be $2^3 = 8$, as we have 3 coins.

$\begin{bmatrix} P \\ Q \\ Q \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$

✓ ✓

when $P = 1$, to satisfy the condition, both Q must be 0
when $P = 0$, to satisfy the condition, both Q must be 1.

Thus, we have 2 out of 8 probability to satisfy the condition.

Prob $= \frac{2}{8} = \frac{1}{4} = 0.25$

(c) If $P = 1$, what is the probability that no quarter produces the same result as the penny, *as a function of Q*?

Solution: $\frac{1}{2^Q}$

For each coin, each flip has two possible results. the total possible results is $2^Q * 2 = 2^{Q+1}$;

when $P=1$, to satisfy the condition, all quarters must be 0
whe $P=0$, to satisfy the condition, all quarters must be 1.

Thus we have 2 out of 2^{Q+1} probability to satisfy the condition

$$\text{prob} = \frac{2}{2^{Q+1}} = \frac{1}{2^Q}$$

(d) Now consider $P = 2$. Given that the first penny produced a different result from all of the quarters, what is the probability that the second penny did as well (as a function of Q)?

Solution: 0.5

For the given condition, there might be 2 possible result of first P_1.

when $P_1 = 1$, we must have all quarters are 0, when $P_1 = 0$, all quarters are 1

$$\begin{bmatrix} P \\ P \\ Q \\ Q \\ \vdots \\ Q \end{bmatrix} = \begin{bmatrix} 1 \\ 1/0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \text{ or } \begin{bmatrix} 0 \\ 1/0 \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

The second penny could have 2 possible results, thus under the first given condition, we have 4 possible results, when $P_1 = 1$, $P_2 = 1$, or $P_1 = 0$, $P_2 = 0$, it satisfies the second condition. Therefore, the probability is

$$\text{prob} = \frac{2}{4} = \frac{1}{2} = 0.5$$

(e) What is the probability that neither of the pennies produced the same result as any of the quarters?

$1/2^{Q+1}$

Solution: $\begin{bmatrix} P \\ P \\ Q \\ Q \\ \vdots \\ Q \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$

There are two possible results to satisfy the condition.

For each possible result, the flip gets 1 or 0 has 0.5 probability.

Therefore, we have $prob = 2 \times \frac{1}{2^2} \times \frac{1}{2^Q} = \frac{1}{2^{Q+1}}$

(f) For $D = 1$, what is the probability that with P pennies and Q quarters, no pair of unlike coins produced the same result, *as a function of P and Q*? This is the probability that these two classes are separable in 1 dimension.

Solution: For $D=1$, to satisfy the condition, there are two possible results, that is

$\begin{bmatrix} P \\ P \\ \vdots \\ P \\ Q \\ Q \\ \vdots \\ Q \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$ or $\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$

For each flip, it has 0.5 probabilities to get 0 or 1.

Therefore, we have total probability:

$$prob = 2 \times \frac{1}{2^P} \times \frac{1}{2^Q} = \frac{1}{2^{P+Q-1}}$$

(40) 2. Now consider the case where each coin is flipped twice ($D = 2$).

(a) Consider $P = 1$ and $Q = 1$. That is, we flip one penny and one quarter two times each. Each coin now has four possible results: $[0,0]$, $[0,1]$, $[1,0]$, and $[1,1]$. What is the probability that the two coins produce different results?

Solution: $3/4 = 0.75$

To solve the problem, we could calculate the probabity that the two coins produce the same results.

For each coin, when we flip twice, we have $2 \times 2 = 4$ possible results, then for two coins, we have $4^2 = 16$ possible results.

For each possible result of P, the second coin get the same result would be 1/4, $\text{prob}_1 = 4 \times \frac{1}{4} \times \frac{1}{4}$, then to solve the problem as given condition, we have

$$\text{prob} = 1 - \text{prob}_1 = 1 - \frac{4}{4 \times 4} = \frac{3}{4} = 0.75$$

(b) For $P = 1$ and $Q = 2$, what is the probability that neither quarter produces the same result as the penny? $9/16$

Solution: There are four possible results for P. and to satisfy the given condition, for each possible result, the 2 quarters must be selected from the left 3 results. So we have $4 \times 3 \times 3$ possible results to satisfy the problem.

Total possible results $= 4^P \times 4^Q = 4^3$. Then we have:

$$\text{prob} = \frac{4 \times 3 \times 3}{4^3} = \frac{9}{16}$$

(c) Finally, for $P = 1$, what is the probability that no quarter produces the same result as the penny, *as a function of Q*?

Solution. There are 4 possible results for P, to satisfy the problem given condition, all the quarters must have flip results selected from the left 3 results. That is, we have 4×3^Q possible results. For total flips, we have $4^{P+Q} = 4^{Q+1}$ possible results. So

$$\text{prob} = \frac{4 \times 3^Q}{4^{1+Q}} = \left(\frac{3}{4}\right)^Q$$

(30) 3. (a) What trend do you see in the probability of class separability as the number of observations[2] increases?

	D	P	Q	Probability of Class Separability
1(a)	1	1	1	1/2
1(b)	1	1	2	1/4
1(c)	1	1	Q	1/2^Q
1(e)	1	2	Q	1/2^(Q+1)
1(f)	1	P	Q	1/2^(P+Q-1)
2(a)	2	1	1	3/4
2(b)	2	1	2	9/16
2(c)	2	1	Q	(3/4)^Q

solution:
We could see from the analysis results as summarized in the table. For both D=1 and D=2, when the number of coins(the number of observations) increases, the probability of class separability decreases exponentially.
- The maximum value for D=1 is when P=1, Q=1, prob_max = 0.5, then following the function of 1/2^(P+Q-1), it will decrease to nearly 0.
- The maximum value for D=2 ,P=1, is when Q=1, prob_max = 0.75, then following the function of (3/4)^Q, it will decrease to nearly 0.
In the listed images of question 4(a), you could also find the trends of probability of class separability decreases when the number of features increases.
We could see that the probability of class separability is decreasing as the nubmer of observations increases.

(b) What trend do you see in the probability of class separability as the number of features[3] increases?

	D	P	Q	Probability of Class Separability
1(a)	1	1	1	1/2
2(a)	2	1	1	3/4
1(b)	1	1	2	1/4
2(b)	2	1	2	9/16
1(c)	1	1	Q	1/2^Q
2(c)	2	1	Q	(3/4)^Q

solution:
We could see from the table analysis, that given P and Q are the same variables, when the number of features(D) increases, the probability of class separability also increases.
- When P=1, Q=1, D: 1->2, the probability increases from 1/2 to 3/4 .
- When P=1, Q=2, D: 1->2, the probability increases from 1/4 to 9/16.
- When P=1, D: 1->2, the probability function increases from 1/2^Q to (3/4)^Q.
In the listed images of question 4(a), you could also find the trends of probability of class separability increases when the number of features increase.
We could see that the probability of class separability is increasing as the number of features increases.

[2] In the distinguishing coins problem, the number of coins (P and Q) is the number of observations.

[3] In the distinguishing coins problem, the number of flips of each coin (D) is the number of features.

[4] Feel free to run your Monte Carlo simulations in the next section before answering this question, if you think that would be helpful.

(c) How do changes in the number of observations and features impact the potential for overfitting?[4]

Solution:
-When we increase the number of observations, the probability of class separability decreases. That is we will have overfitting problem when the number of observations increase to a certain level. Because it is hard to separate the class, our model becomes overfitting.
-When we increase the number of features, the probability of class separability increases. That is when the number of observation is constant, if we increase our features, we get larger probability of class separability. That is we have greater chance to separate our dataset class. Thus the overfitting problem is reduced.

In a word, increasing the number of observations will intensify overfitting problem, while increasing the number of features will lighten overfitting problem.

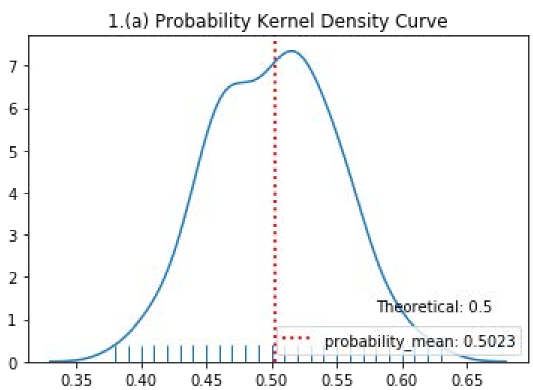
Monte Carlo Simulation

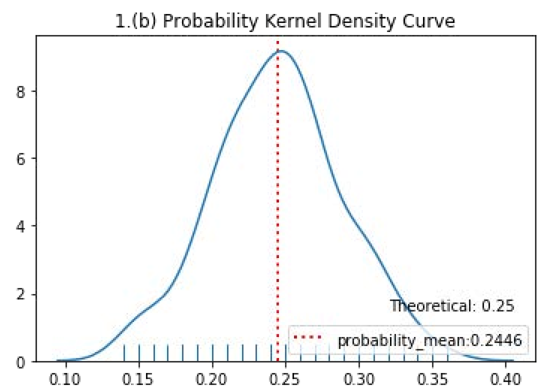
Implement a Monte Carlo simulation of this coin-flipping problem to verify your theoretical analyses of this coin separating problem.

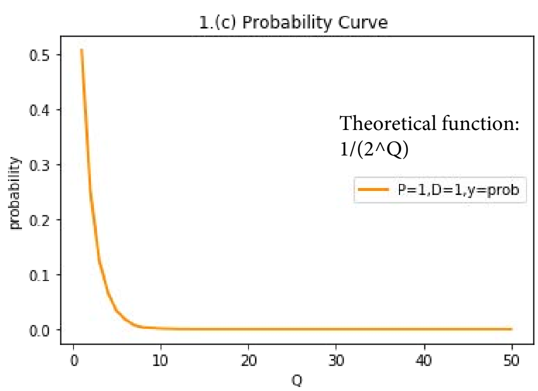
(60) 4. (a) Experimentally verify all the probabilities you theoretically calculated.

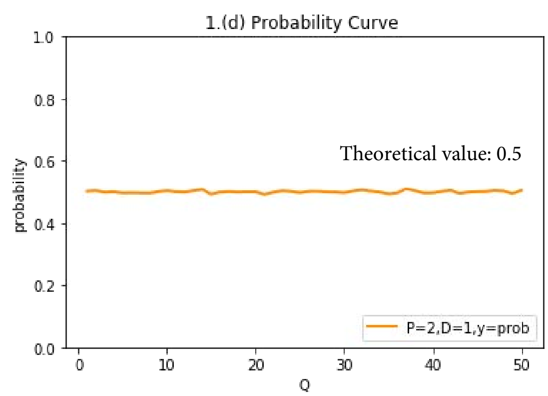
(b) Estimate the probability of class separability for $P = Q = 5$ and $D = 5$.

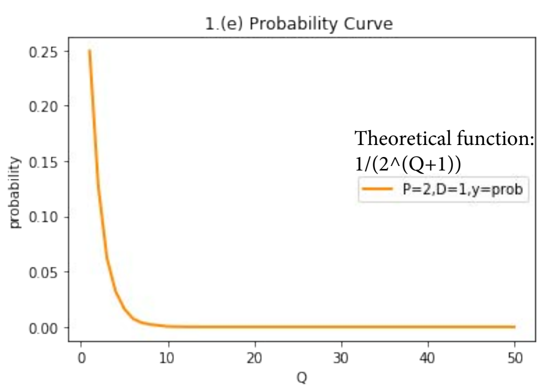
(a) Please see the attached pictures:

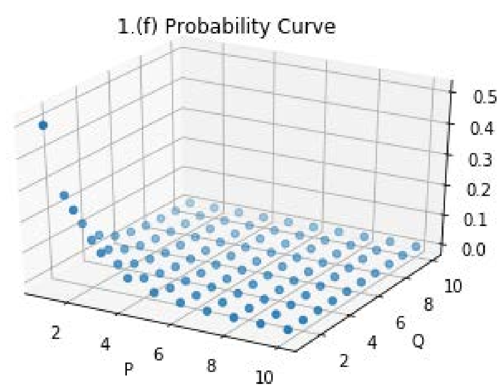




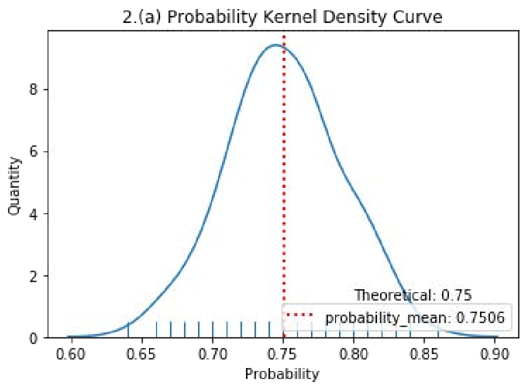


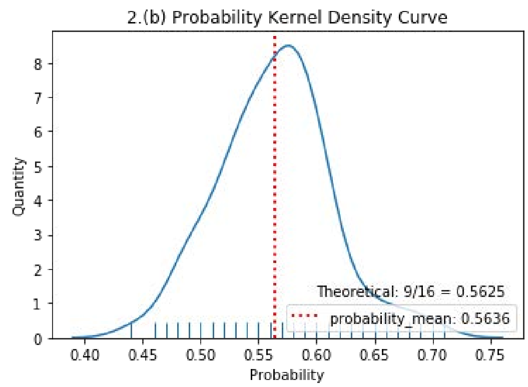


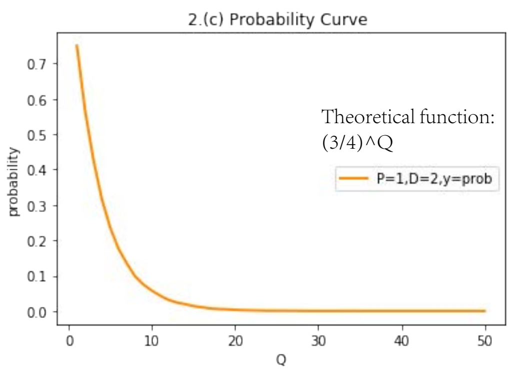




Theoretical function: 1/(2^(P+Q-1))







(b) Solution: Please see attached image as simulation result.
The simulated probability mean is around 0.4510. All code files please see the other attached file.