# Machine Learning, test 1

Name: .....

Student Number: .....

## Part 1, Linear regression

### Q1. Choose one correct statement about linear regression

**A. To learn about the linear functional dependency between the input columns and output column in the data table**

B. To classify the output given one input

C. To model the probability function of output given the input

D. To solve the binary classification problem

**Answer: A. The linear model is for regression, not for classification, nor for probability modeling**

### Q2. What is the cost function to learn the linear model y=wx?

A. $J(w)=(wx_1+y_1)^2+(wx_2+y_2)^2+..+(wx_n+y_n)^2$

B. $J(w)=(wx_1-y_1)^2+(wx_2-y_2)^2+..+(wx_n-y_n)^2$

C. $J(w)=\sum_{j=1..n}(wx_j+y_j)^2$

D. $J(w)=\sum_{j=1..n}(wx_j-y_j)$

**Answer: B. the cost must be sum of positive differences (absolute-different or squared difference)**

### Q3. What is the simplest method to learn the parameter w of the linear model?

A. Simulated Annealing

B. SMO

C. Gradient Ascending

D. SVM

**Answer: C. Gradient Descending is the simplest**

### Q4. Choose the dataset which a linear model can fit in well

| A | | B | | C | | D | |
|---|---|---|---|---|---|---|---|
| 1 | 0 | 1 | 500 | 1 | 0.002718282 | 1 | 0.540302306 |
| 2 | 0.301029996 | 2 | 1001 | 2 | 0.007389056 | 2 | -0.416146837 |
| 3 | 0.477121255 | 3 | 1500 | 3 | 0.020085537 | 3 | -0.989992497 |
| 4 | 0.602059991 | 4 | 2000 | 4 | 0.05459815 | 4 | -0.653643621 |
| 5 | 0.698970004 | 5 | 2500 | 5 | 0.148413159 | 5 | 0.283662185 |
| 6 | 0.77815125 | 6 | 3001 | 6 | 0.403428793 | 6 | 0.960170287 |
| 7 | 0.84509804 | 7 | 3500 | 7 | 1.096633158 | 7 | 0.753902254 |
| 8 | 0.903089987 | 8 | 4000 | 8 | 2.980957987 | 8 | -0.145500034 |
| 9 | 0.954242509 | 9 | 4500 | 9 | 8.103083928 | 9 | -0.911130262 |
| 10 | 1 | 10 | 5001 | 10 | 22.02646579 | 10 | -0.839071529 |

**Answer: B is almost linear**

## Part 2, Logistic regression

### Q5. Which of the following function is the decision function for logistic regression for binary classification case?

A. $y = \begin{cases} face & p(face \mid x) = p(nonface \mid x) \\ nonface & otherwise \end{cases}$

B. $y = \begin{cases} face & p(face \mid x) = 1 \\ nonface & otherwise \end{cases}$

C. $y = \begin{cases} face & w_{face} * x > 1 \\ nonface & otherwise \end{cases}$ where $p(face \mid x) = \frac{1}{1 + \exp(w_{face} * x)}$

D. $y = \begin{cases} face & w_{face} * x < 0 \\ nonface & otherwise \end{cases}$ where $p(face \mid x) = \frac{1}{1 + \exp(w_{face} * x)}$

**Answer: D. We have w*x < 0 then exp(w*x) < 1 then p(face|x) = 1/(1+exp(w*x)) > 0.5**

### Q6. What is the simplest method to learn the parameter w of the logistic model?

A. Simulated Annealing

B. SMO

C. SVM

D. Gradient Ascending

**Answer: D. Gradient Descending is the simplest**

### Q7. When we want to categorize three classes (sky, grass, tree) from background, how many probability functions do we have to design with multinomial logistic regression?

A. 1 B. 2 C. 3 D. 4

**Answer: C. For 4 class, we need to learn 3 functions, the last is the reference/pivot class**

## Part 3, Mixture model, kmeans, EM

### Q8. Select the correct statement about mixture model?

A. We can combine two probability distribution function p1(x) and p2(x) with any real number w1 and w2 to create a mixture model.

B. p1(x) and p2(x) must be both gaussian distribution or both uniform distribution in order to mix. We cannot mix gausian distribution with uniform distribution.

C. We need two positive numbers w1 and w2 in which w1+w2 = 1 in order to obtain the mixture model.

D. We can mix any functions p1(x) and p2(x) using two positive numbers w1 and w2 with w1+w2 = 1 to create the mixture model.

**Answer: C we can ignore general function (D) or general number (A).**

### Q9. Select the correct statement about Gaussian mixture model (GMM)?

A. GMM is a mixture model, the components of which are uniform distributions.

B. GMM can be used to model classes which can be divided into elliptic subsets.

C. GMM can be gaussian model when the number of subsets (centers) equals to 2.

D. We can replace GMM with kmeans because GMM is more general than kmeans.

**Answer: B. Elliptic subsets is the one. GMM cannot replace kmeans. GMM can be gaussian when k=1**

### Q10-13. Match the left and the right parts to form the correct statements about kmeans procedure

| | |
|---|---|
| In labeling step, ....(10)... | A. we have to perform cross validation. |
| In averaging step, ....(11)... | B. we can randomize the labels or the centers. |
| In initializing step, ....(12)... | C. we can compute the centers (means) from the labels in the previous step. |
| In order to choose parameter k, ....(13)... | D. we can compute the labels from the centers (means) in the previous step. |

**Answer: 10-D, 11-C, 12-B, 13-A**

### Q14-16. Match the left and the right parts to form the correct statements about EM procedure

| | |
|---|---|
| In M step, ....(14)... | A. we have to perform cross validation. |
| In E step, ....(15)... | B. we need to randomize the centers. |
| In initializing step, ....(16)... | C. we need to normalize the labels |
| | D. we need to average all the samples with the labels for each center. |

**Answer: 14-D, 15-C, 16-B**