

Practice exam questions

1. Classifying problems

- Explain the difference between supervised and unsupervised learning.
- Explain the difference between classification and regression.

For each of the problem descriptions below, indicate what *type* of machine learning algorithm would be most suited to solving it. So, would you use a supervised or unsupervised approach for the problem, and specifically, would you use regression, classification or clustering. Motivate your answers.

- Predict the number of millimeters of rain that will fall tomorrow, based on attributes like humidity and atmospheric pressure, using meteorological records of these attributes and the measured rainfall from previous days.
- Using patient data from earlier studies, where several patient were monitored for 6 months for blood values such as glucose and cholesterol levels, and whether they had a heart attack in that time period, predict if a new patient is likely to have a heart attack based on their current blood values.

2. Data sets

- Explain the difference between training and validation data, and how both are used by a supervised learning algorithm.
- Explain the difference between validation and testing data, and why these two parts of the data should not overlap.

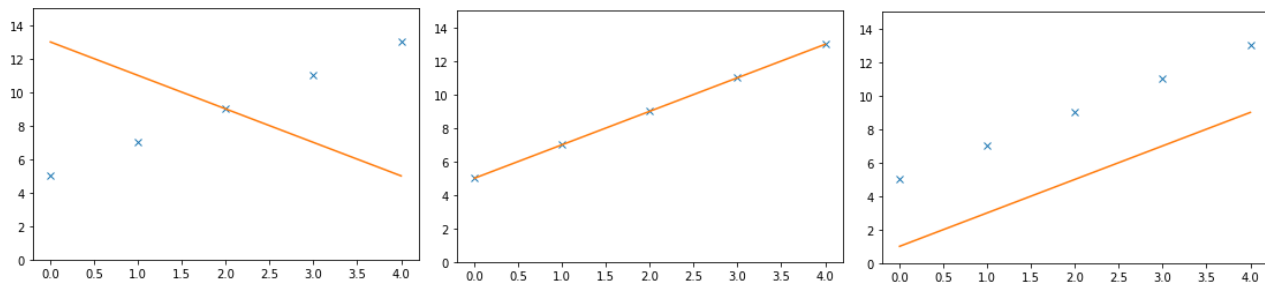
3. k-Nearest Neighbours

- For which value of k is k-NN equal to the nearest neighbour algorithm?
- Explain why you would always use an uneven value for k specifically when doing binary classification.

4. Naive Bayes

- Naive Bayes uses Bayes' rule to rewrite the posterior probability in terms of the prior and likelihood. Explain why this rewrite is done, instead of just estimating the posterior probability directly.
- The estimations of the likelihood in this algorithm are considered *naive*, at least, that is where the algorithm gets the other half of its name. Explain why this computation is considered a naive.

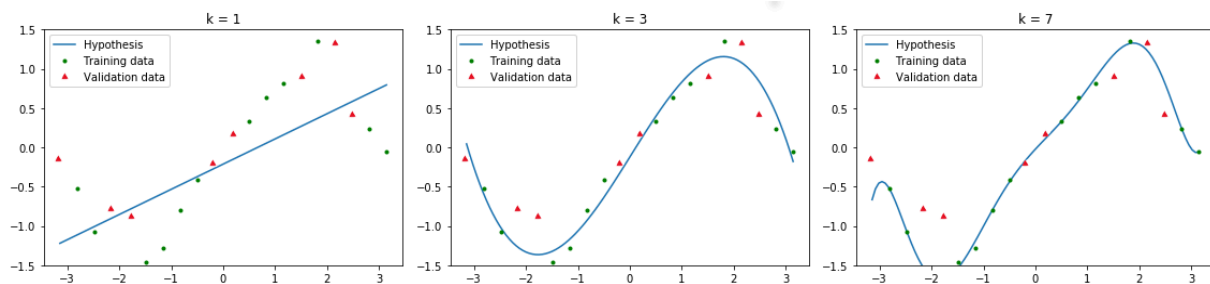
5. Linear Regression



The figure above shows 3 different linear regression models, for the same 5 data points.

- Order these 3 linear models from lowest to highest cost for the 5 plotted data points. No need to compute the cost exactly, but motivate your answer for each graph.
- Explain why the point where the gradient of the cost is exactly 0 in both directions, is the also best possible value of a and b for the data set.
- Explain what the learning rate is and what it does in gradient descent.

6. Polynomial Fit



Consider the univariate regression data above. The green dots are the training samples, the red triangles are the validation samples and the blue line represents the learned hypothesis. The each model is learned using polynomial regression, using a different polynomial degree, which is indicated above the each plot by the k parameter.

- Which of these plots ($k=1, 3$ and 7) results in the lowest training cost? And which plot results in the lowest validation cost? Explain your answers.
- For each of the three plots, indicate whether the model underfits, overfits or fits well. For each of your three answers, explain what parts of the plot indicate that type of fit.
- For each plot separately, answers the following question: What are the dimensions of the training data matrix after the polynomial feature expansion is applied?

(num of training data, $k+1$)
= $k+1$ RMB!

7. Polynomial Cost

The following equations define the polynomial feature expansion, prediction and cost of polynomial regression model.

$$X = \begin{bmatrix} x^{(1)} & (x^{(1)})^2 & (x^{(1)})^3 & \dots & (x^{(1)})^n \\ x^{(2)} & (x^{(2)})^2 & (x^{(2)})^3 & \dots & (x^{(2)})^n \\ x^{(3)} & (x^{(3)})^2 & (x^{(3)})^3 & \dots & (x^{(3)})^n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x^{(m)} & (x^{(m)})^2 & (x^{(m)})^3 & \dots & (x^{(m)})^n \end{bmatrix}$$

$$\hat{\mathbf{y}} = X\mathbf{w} \oplus b$$

$$J(\mathbf{w}, b) = \frac{1}{2m} \sum_{i=1}^m (\hat{y}^{(i)} - y^{(i)})^2$$

You'll be applying this model to a simple data set consisting of 3 points and using the following starting values for \mathbf{w} and b .

$$\mathbf{x} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

$$\mathbf{y} = \begin{bmatrix} 2 \\ 5 \\ 3 \end{bmatrix}$$

$$\mathbf{w} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

$$b = 3$$

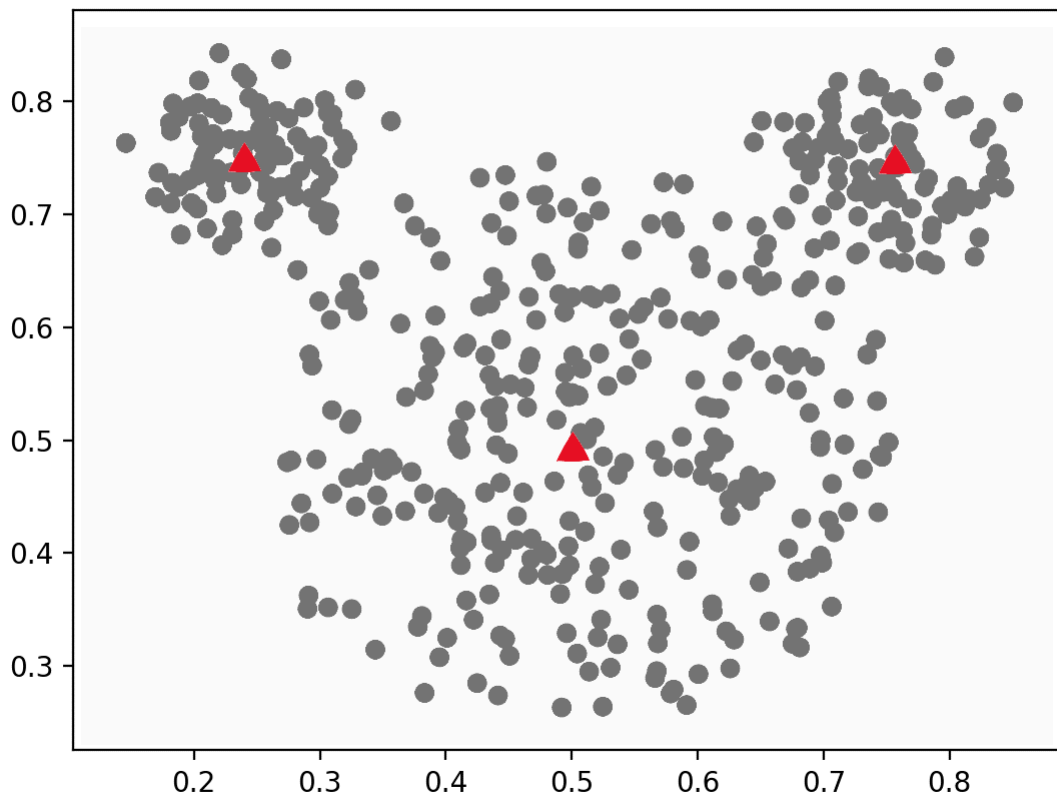
- Write out the expansion matrix X for a quadratic (2nd degree) model.
- Write out the prediction vector $\hat{\mathbf{y}}$ using the provided starting values for \mathbf{w} and b .
- Compute the cost for the prediction vector $\hat{\mathbf{y}}$ from the previous question.
Note: The real exam might also contain computation questions, so you can (and should) bring a calculator to the exam.
- Draw the 3 data points in a plot and draw in what you think the final fitted quadratic model would look like *after* performing gradient descent. Your line for the model does not need to be exact, but the rough shape should match the type of the model.
Note: The real exam might also contain questions that require you to draw plots, so you can (and should) bring a pencil and ruler to the exam.

$$X = \begin{bmatrix} 1 & 1 \\ 2 & 4 \\ 3 & 9 \end{bmatrix}$$

$$\hat{\mathbf{y}} = X\mathbf{w} \oplus b$$

$$= \begin{bmatrix} 4 \\ 12 \\ 24 \end{bmatrix}$$

8. k-Means



The figure above shows a data set that is being clustered using the k-means algorithm. The data is indicated by dots, and the current means are indicated by red triangles.

- Run a single iteration of the k-means algorithm on this data set, doing both steps (no need to write out all results here). Does the algorithm converge after this iteration? Explain your answer.
- Name two methods you could use to determine what value of k you should use for k-means.

elbow method
 plot cost against k .
 find where improvement drops.
 use downstream use.
 the one