

# 1 Week 7 Unsupervised Learning

## 1.1 K-means Clustering

- In this exercise, you will implement the K-means algorithm and use it for image compression.
- You will first start on an example 2D dataset that will help you gain an intuition of how the K-means algorithm works. After that, you will use the K-means algorithm for image compression by reducing the number of colors that occur in an image to only those that are most common in that image.
- You will be using `ex7.m` for this part of the exercise.

## 1.2 Implementing K-means

- The K-means algorithm is a method to automatically cluster similar data examples together.
- The intuition behind K-means is an iterative procedure that starts by guessing the initial centroids, and then refines this guess by repeatedly assigning examples to their closest centroids and then recomputing the centroids based on the assignments.
- The inner-loop of the algorithm repeatedly carries out two steps:
  - (i) Assigning each training example to its closest centroid
  - (ii) Recomputing the mean of each centroid using the points assigned to it.
- The K-means algorithm will always converge to some final set of means for the centroids. Note that the converged solution may not always be ideal and depends on the initial setting of the centroids.
- Therefore, in practice the K-means algorithm is usually run a few times with different random initializations.
- One way to choose between these different solutions from different random initializations is to choose the one with the lowest cost function value (distortion).
- Random initialization The initial assignments of centroids for the example dataset in `ex7.m` were designed so that you will see the same figure as in Figure 1.
- In practice, a good strategy for initializing the centroids is to select random examples from the training set.

### Question 1.

For which of the following tasks might K-means clustering be a suitable algorithm? Select all that apply.

Question 2.

Suppose we have three cluster centroids  $\mu_1=[12]^T$ ,  $\mu_2=[-30]^T$  and  $\mu_3=[42]^T$ . Furthermore, we have a training example  $x(i)=[31]^T$ . After a cluster assignment step, what will  $c^{(i)}$  be?

Question 3.

K-means is an iterative algorithm, and two of the following steps are repeatedly carried out in its inner-loop. Which two?

Question 4.

Suppose you have an unlabeled dataset  $\{x(1), \dots, x(m)\}$ . You run K-means with 50 different random initializations, and obtain 50 different clusterings of the data. What is the recommended way for choosing which one of these 50 clusterings to use?

Question 5.

Which of the following statements are true? Select all that apply.

## Anomaly Detection

- Suppose you are developing an anomaly detection system to catch manufacturing defects in airplane engines.

- Your model uses

$$p(x) = \prod_{j=1}^n p(x_j; \mu_j, \sigma_j^2)$$

- You have two features  $x_1 = \textit{vibration intensity}$ , and  $x_2 = \textit{heat generated}$ .
- Both  $x_1$  and  $x_2$  take on values between 0 and 1 (and are strictly greater than 0), and for most "normal" engines you expect that  $x_1 \approx x_2$ .
- One of the suspected anomalies is that a flawed engine may vibrate very intensely even without generating much heat (large  $x_1$ , small  $x_2$ ), even though the particular values of  $x_1$  and  $x_2$  may not fall outside their typical ranges of values.
- What additional feature  $x_3$  should you create to capture these types of anomalies:

### Solution Options

- $x_3 = x_1 + x_2$  This could take on large or small values for both normal and anomalous examples, so it is not a good feature.

## 2 Week 9 Quiz. Anomaly Detection

### Question 1.

For which of the following problems would **anomaly detection** be a suitable algorithm?

- From a large set of hospital patient records, predict which patients have a particular disease (say, the flu).
- From a large set of primary care patient records, identify individuals who might have unusual health conditions.
- CORRECT Given data from credit card transactions, classify each transaction according to type of purchase (for example: food, transportation, clothing).
- In a computer chip fabrication plant, identify microchips that might be defective.

### Question 2. Variant A

Suppose you have trained an anomaly detection system that flags anomalies when  $p(x)$  is less than  $\epsilon$ , and you find on the cross-validation set that it has too many false negatives (failing to flag a lot of anomalies). What should you do?

- Increase  $\epsilon$  [CORRECT]
- Decrease  $\epsilon$

### Question 2. Variant B

Suppose you have trained an anomaly detection system for fraud detection, and your system that flags anomalies when  $p(x)$  is less than  $\epsilon$ , and you find on the cross-validation set that it mis-flagging far too many good transactions as fraudulent. What should you do?

- Increase  $\epsilon$
- Decrease  $\epsilon$  [CORRECT]

### Question 3.

Suppose you are developing an anomaly detection system to catch manufacturing defects in airplane engines. Your model uses

You have two features  $x_1$  = vibration intensity, and  $x_2$  = heat generated.

Both  $x_1$  and  $x_2$  take on values between 0 and 1 (and are strictly greater than 0), and for most "normal" engines you expect that  $x_1 \approx x_2$ .

One of the suspected anomalies is that a flawed engine may vibrate very intensely even without generating much heat (large  $x_1$ , small  $x_2$ ), even though the particular values of  $x_1$  and  $x_2$  may not fall outside their typical ranges of values. What additional feature  $x_3$  should you create to capture these types of anomalies:

- $x_3 = x_1/x_2$  [CORRECT]
- $x_3 = x_2 \text{ times } x_2$
- $x_3 = x_1 \text{ times } x_2$
- $x_3 = x_1 + x_2$

### Question 4.

Which of the following are true? Check all that apply.

- When evaluating an anomaly detection algorithm on the cross validation set (containing some positive and some negative examples), classification accuracy is usually a good evaluation metric to use.
- In anomaly detection, we fit a model  $p(x)$  to a set of negative ( $y=0$ ) examples, without using any positive examples we may have collected of previously observed anomalies.
- **CORRECT** When developing an anomaly detection system, it is often useful to select an appropriate numerical performance metric to evaluate the effectiveness of the learning algorithm.
- In a typical anomaly detection setting, we have a large number of anomalous examples, and a relatively small number of normal/non-anomalous examples.

### Question 5.

You have a 1-D dataset  $\{x(1), \dots, x(m)\}$  and you want to detect outliers in the dataset. You first plot the dataset and it looks like this:

Suppose you fit the gaussian distribution parameters  $\mu_1$  and  $\sigma_1^2$  to this dataset. Which of the following values for  $\mu_1$  and  $\sigma_1^2$  might you get?

- $\mu_1=-3, \sigma_1^2=4$
- $\mu_1=-6, \sigma_1^2=4$
- $\mu_1=-3, \sigma_1^2=2$
- $\mu_1=-6, \sigma_1^2=2$

## 3 Week 9 Quiz. Recommender Systems

Information Filtering System that attempts to recommend information items likely to be of interest to a user.

### Commonly used algorithms

- $k$ -means clustering
- Pearson's Rho
- Collaborative Filtering

Collaborative Filtering is the process of filtering for information or patterns using collaboration among multiple agents.

Applications: online news aggregation or similar items of clothings

best approached by other methods - prediction

Collaborative Filtering Gradient

$$\frac{\partial J}{\partial X_k^{(i)}} = \sum \mathbb{I} \theta_k^{(j)}$$
$$\frac{\partial J}{\partial \theta_k^{(i)}} = \sum \mathbb{I} X_k^{(j)}$$

No regularization applied

Anomaly Detection Gaussian Distribution Estimate Gaussian Distribution

For  $n$  features of  $X$ , compute the mean and variance for each feature

Selecting Threshold of  $\epsilon$ .

Implement an algorithm to select the threshold  $\epsilon$  using an  $F_i$  score on a cross validation set.

$P(X) < \epsilon$  is considered to be an anomaly.

### 3.1 $F_1$ Score