Discussion Session Problems 7

3/5/2026

1 K-Means Clustering

Question 1. K-means is not deterministic.

a) True

b) False

Question 2. K-means works:

a) better with data where clusters have similar variance

b) better with data where each cluster has its own variance

c) equally as well with both

Question 3. What do you do with the knowledge that K-means can get stuck in a local minimum?

Question 4. It would be helpful to normalise all features in the dataset to have the same variance (unit variance) before running K-means.

a) True

b) False

2 K-Nearest Neighbours

Question 5. K-Nearest Neighbours (KNN), a model that assigns a data point to the class of its K nearest neighbours, will often work *better* than softmax classification for highly nonlinear data.

a) True

b) False

Question 6. A K-Nearest Neighbours model will often work *better* than softmax classification for data with many features.

a) True

b) False

3 Principal Component Analysis (PCA)

Question 7. If we run PCA on a dataset it forms linear combinations of the features, allowing us to run linear regression with weights of just $+1$ and -1.

a) True

b) False

Question 8. What would happen if you run PCA without normalising the dataset?

Question 9. What does it mean if you get zero eigenvalues in the correlation matrix when running PCA?

Question 10. A principal components analysis was run and the following eigenvalue results were obtained:

$$\lambda = [2.731,\ 2.218,\ 1.442,\ 0.009,\ 0.00183,\ 0.00085]$$

How many components would you retain?

a) 2

b) 3

c) 4

d) 5

4 Anomaly Detection

Question 11. When do you decide to employ an unsupervised anomaly detection algorithm instead of a supervised learning algorithm?

5 Recommender Systems

Question 12. Collaborative filtering is most useful in cases where users' tastes can change and evolve over time.

a) True

b) False

Question 13. Embeddings and Similarity. A collaborative filtering model has been trained *solely* on user star ratings — it has no access to movie metadata such as genre, director, cast, or plot description. After training, two movies are found to have very similar learned embedding vectors $\mathbf{x}^{(i)}$ and $\mathbf{x}^{(k)}$. What is the most accurate interpretation?

a) The two movies were directed by the same person

b) The two movies have been rated by the same users

c) The two movies share similar latent characteristics as inferred from user rating patterns

d) The two movies belong to the same genre as specified in their metadata

Question 14. Content-Based Filtering — Worked Prediction. A platform has three movies with known feature vectors capturing [action, romance] levels on a 0–1 scale. User A has learned weights $\mathbf{w}^{(A)} = [0.9,\ 0.1]$ and bias $b^{(A)} = 0.5$.

(a) Using the prediction formula $\hat{y}^{(i,j)} = (\mathbf{w}^{(j)})^\top \mathbf{x}^{(i)} + b^{(j)}$, compute the predicted rating for each movie below. Which movie does the model recommend to User A, and does it make intuitive sense?

Movie	$\mathbf{x}^{(i)}$	$\hat{y}^{(A,i)}$
Movie 1	[0.8, 0.1]	_____
Movie 2	[0.2, 0.9]	_____
Movie 3	[0.7, 0.3]	_____

(b) A brand new movie is added with no ratings yet. Can content-based filtering still recommend it? Can collaborative filtering? Explain the difference in one or two sentences — this is the **cold-start problem**.

Question 15. Mean Normalization. The table below shows ratings (1–5 stars). "?" means not yet rated. User C has no ratings at all.

	User A	User B	User C	μ_i
Movie 1	4	5	?	_____
Movie 2	2	?	?	_____
Movie 3	5	4	?	_____

(a) Compute the per-item mean $\mu_i = \frac{\sum_j r(i,j)\, y^{(i,j)}}{\sum_j r(i,j)}$ for each movie. Then compute the normalized rating $\tilde{y}^{(i,j)} = y^{(i,j)} - \mu_i$ for every observed entry.

(b) Without mean normalization, what rating would the model predict for User C on Movie 1 (whose $\mathbf{w}$ and b are initialized to zero)? With mean normalization the final prediction is $\hat{y}^{(i,j)} = (\mathbf{w}^{(j)})^\top \mathbf{x}^{(i)} + b^{(j)} + \mu_i$. What does this predict for User C, and why is it more sensible?