Machine Learning

V. Adamchik | CSCI 567 | Fall 2019

Discussion Set 10 | University of Southern California

GMM, EM

Problem 1

Which of the following statements of the Expectation-Maximization (EM) algorithm is true?

(A) Before running the EM algorithm, we need to choose the step size.

(B) EM always converges to the global optimum of the likelihood.

(C) In EM, the lower-bound for the log-likelihood function we maximize is always non-concave.

(D) None of the above.

Problem 2

Which of the following statements of Gaussian Mixture Model (GMM) is true?

(A) GMM is a non-parametric method that all the training samples need to be stored.

(B) GMM is a probabilistic model that can be used to explain how data is generated.

(C) When learning a GMM, the labels of the samples are available.

(D) None of the above

Problem 3

Which of the following statements of the Expectation-Maximization (EM) algorithm is correct?

(A) The EM algorithm maximizes the expectation of latent variables $E[z_n]$.

(B) In the expectation (E) step, we can use an arbitrary distribution $q_n(z_n)$ to maximize the lower bound of the log-likelihood $P(\Theta)$.

(C) In the maximization (M) step, the objective likelihood is guaranteed to be increased by updating the model parameters (if not converged).

(D) None of above.

Problem 4

Which of the following statements of Gaussian Mixture Model (GMM) is correct?

(A) The parameters of a GMM can be learned via maximum-likelihood estimation (MLE).

(B) Gradient descent cannot be used to learn a GMM. not changed since this is the incorrect answer

(C) GMM is a supervised learning method.

(D) None of above.

Problem 5

Which of the following statements of GMM and K-means is correct?

(A) Given a set of data points and a fixed number of clusters K, applying GMM and K-means will always result in same cluster centroids.

(B) Given a set of data points and a fixed number of clusters K, applying GMM and K-means will always result in different cluster centroids.

(C) Given a learned GMM, we assign a data point to a cluster if the distance from the data point to its centroid is the smallest.

(D) Given a learned K-means model, we assign a data point to a cluster if the distance from the data point to its centroid is the smallest.

Problem 6

Which of the following is not correct?

(A) We can generate novel samples of data from a Gaussian Mixture Model.

(B) Naive Bayes classier (both continuous or discrete input) is a linear classier.

(C) Parameters of Logistic Regression can be derived in closed form from parameters of Naive Bayes classifier.

(D) Generative models can perform better when there is less labelled data for training compared to discriminative models.

Problem 7

Let x be a one-dimensional random variable distributed according to a mixture of two Gaussian distributions

$$P(x) = \omega_1 N\left(x\middle|\mu_1, \sigma_1^2\right) + \omega_2 N(x|\mu_2, \sigma_2^2)$$

M-step: derive the updates for the parameters ω_k and σ_k^2.