Exercise 16 : Mock-up exam

Exercise 16-1 :

Which of the following statements are true about the Expectation Maximization (EM) and k-means clustering algorithms ?

(a) The EM algorithm can learn covariances between feature dimensions while the $k-$mean algorithm cannot.

(b) After being trained with two different choices of k, the k-means algorithm can provide a preference score, while EM cannot.

(c) While it is possible to choose the optimal k after training EM once on a data set, it is not possible for $k-$means.

(d) While the Mac-Queen version of the $k-$means output is dependent on the processing order of the data points, the EM output is not.

(e) While the Lloyd-Forgy version of the $k-$means output is dependent on the processing order of the data points, the EM output is not.

(f) When trained on the same cluster count k, Expectation Maximization has more parameters to fit than k-means.

Exercise 16-2 :

Given a supervised predictor $f(x) = W_3^T \sigma(W_2^T \sigma(W_1^T x))$, where $\sigma(u) = tanh(u)$ and the superscript T denotes matrix transpose, which takes x as an input feature vector and predicts the output label y, a training set $\mathcal{D}_{train} = \{(x_n, y_n) | n = 1, \ldots, N\}$ and a loss function

$$\mathbb{L}(W_1, W_2, W_3, \mathcal{D}_{train}) = \frac{1}{N} \sum_{n=1}^{N} \Big(y_n - f_{W1,W2,W3}(x_n) \Big)^2.$$

Which of the following statements are correct about the optimization problem below?

$$\underset{W1,W2,W3}{\operatorname{argmin}} \; \mathbb{L}(W_1, W_2, W_3, \mathcal{D}_{train})$$

(a) Solving $\nabla_{W1,W2,W3} \mathbb{L}(W_1, W_2, W_3, \mathcal{D}_{train}) = 0$ for $W1, W2, W3$ finds the unique global minimum for $\mathbb{L}$.

(b) Solving $\nabla_{W1,W2,W3} \mathbb{L}(W_1, W_2, W_3, \mathcal{D}_{train}) = 0$ for $W1, W2, W3$ finds one of the many global minima for $\mathbb{L}$.

(c) Initializing the weights (W_1, W_2, W_3) to random values and updating them according to

$$(W_1, W_2, W_3) := (W_1, W_2, W_3) - \alpha \nabla_{W_1,W_2,W_3} \mathbb{L}(W_1, W_2, W_3, \mathcal{D}_{train})$$

for learning rate $\alpha > 0$ until convergence finds the unique global minimum for $\mathbb{L}$.

(d) Initializing the weights (W_1, W_2, W_3) to random values and updating them according to

$$(W_1, W_2, W_3) := (W_1, W_2, W_3) + \alpha \nabla_{W_1,W_2,W_3} \mathbb{L}(W_1, W_2, W_3, \mathcal{D}_{train})$$

for learning rate $\alpha > 0$ until convergence finds a local minimum for $\mathbb{L}$.

(e) Initializing the weights (W_1, W_2, W_3) to random values and updating them according to

$$(W_1, W_2, W_3) := (W_1, W_2, W_3) - \alpha \nabla_{W_1,W_2,W_3} \mathbb{L}(W_1, W_2, W_3, \mathcal{D}_{train})$$

for learning rate $\alpha > 0$ until convergence finds a local minimum for $\mathbb{L}$.

(f) Solving $\nabla_{W1,W2,W3} \mathbb{L}(W_1, W_2, W_3, \mathcal{D}_{train}) = 0$ for $W1, W2, W3$ analytically gives the unique global minimum for $\mathbb{L}$ when $\sigma(u) = u$.

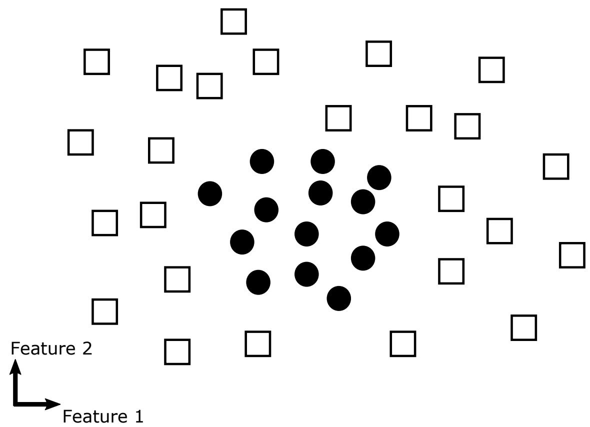
(g) Solving $\nabla_{W1,W2,W3} \mathbb{L}(W_1, W_2, W_3, \mathcal{D}_{train}) = 0$ for $W1, W2, W3$ analytically gives the unique global minimum for $\mathbb{L}$ when $\sigma(u) = \beta u$ for some $\beta > 0$.

(h) Solving $\nabla_{W1,W2,W3} \mathbb{L}(W_1, W_2, W_3, \mathcal{D}_{train}) = 0$ for $W1, W2, W3$ analytically gives the unique global minimum for $\mathbb{L}$ when

$$\sigma(u) = 1/(1 + e^{-u}).$$

Exercise 16-3 :

Consider the binary classification problem below



Which of the below classifiers have the potential to give zero prediction error ?

(a) Neural net with two hidden layers and activation function $\sigma(u) = max(0, u)$.

(b) Neural net with seven hidden layers and activation function $\sigma(u) = u$.

(c) SVM with linear kernel

(d) SVM with squared exponential kernel

(e) Linear regression

(f) Decision tree with axis-aligned splits max depth 2

(g) Decision tree with axis-aligned splits, unlimited max dept