TD 6: Exercises from previous exams

▶ Exercise 1: Logistic regression (credits to EPFL CS-433)

Consider the logistic regression loss $\mathcal{L} : \mathbb{R}^p \to \mathbb{R}$ for a binary classification task with data $(\mathbf{x}_i, y_i) \in \mathbb{R}^p \times \{0, 1\}$:

$$\mathcal{L}(\boldsymbol{\beta}) = \frac{1}{N} \sum_{i=1}^{N} \left(\log \left(1 + e^{\mathbf{x}_i^\top \boldsymbol{\beta}}\right) - y_i \mathbf{x}_i^\top \boldsymbol{\beta} \right)$$

Which of the following is a gradient of the loss $\mathcal{L}$?

(A) $\nabla \mathcal{L}(\boldsymbol{\beta}) = \dfrac{1}{N} \sum_{i=1}^{N} \left(\mathbf{x}_i \dfrac{e^{\mathbf{x}_i^\top \boldsymbol{\beta}}}{1 + e^{\mathbf{x}_i^\top \boldsymbol{\beta}}} - y_i \mathbf{x}_i^\top \boldsymbol{\beta} \right)$

(B) $\nabla \mathcal{L}(\boldsymbol{\beta}) = \dfrac{1}{N} \sum_{i=1}^{N} \mathbf{x}_i \left(y_i - \dfrac{e^{\mathbf{x}_i^\top \boldsymbol{\beta}}}{1 + e^{\mathbf{x}_i^\top \boldsymbol{\beta}}} \right)$

(C) $\nabla \mathcal{L}(\boldsymbol{\beta}) = \dfrac{1}{N} \sum_{i=1}^{N} \mathbf{x}_i \left(\dfrac{1}{1 + e^{\mathbf{x}_i^\top \boldsymbol{\beta}}} - y_i \right)$

(D) None of the above.

▶ Exercise 2: Logistic regression (credits to Berkeley CS-189)

Logistic regression

(A) Assumes that the distribution of the data is Gaussian

(B) Minimizes a convex cost function

(C) Has a closed-form solution

(D) None of the above

▶ Exercise 3 (credits to EPFL CS-433)

Binary logistic regression assumes

(A) Linear relationship between the input variables.

(B) Linear relationship between the observations.

(C) Linear relationship between the input variables and the inverse sigmoid of the probability of the event that the outcome $Y = 1$.

(D) Linear relationship between the input variables and the probability of the event that the outcome $Y = 1$.

▶ Exercise 4: Two gaussians (credits to Berkeley CS-189)

Say we have two 2-dimensional Gaussian distributions representing two different classes. Which of the following conditions will result in a linear decision boundary

(A) Same covariance matrix for both classes.

(B) Same mean for both classes.

(C) Different covariance matrix for each class.

(D) None of the above.

▶ Exercise 5: Gaussian discriminant analysis (credits to Berkeley CS-189)

Which of the following are true about two-class Gaussian discriminant analysis? Assume you have estimated the parameters $\hat{\boldsymbol{\mu}}_1, \hat{\boldsymbol{\Sigma}}_1, \hat{\pi}_1$ for class 1 and $\hat{\boldsymbol{\mu}}_0, \hat{\boldsymbol{\Sigma}}_0, \hat{\pi}_0$ for class 0.

(A) If $\hat{\boldsymbol{\mu}}_1 = \hat{\boldsymbol{\mu}}_0$ and $\hat{\pi}_1 = \hat{\pi}_0$, then the LDA and QDA classifiers are identical.

(B) If $\hat{\boldsymbol{\Sigma}}_1 = \boldsymbol{I}$ and $\hat{\boldsymbol{\Sigma}}_0 = 5\boldsymbol{I}$, then the LDA and QDA classifiers are identical.

(C) If $\hat{\boldsymbol{\Sigma}}_1 = \hat{\boldsymbol{\Sigma}}_0$, $\hat{\pi}_0 = 1/6$ and $\hat{\pi}_1 = 5/6$, then the LDA and QDA classifiers are identical.

(D) None of the above.

▶ Exercise 6 (credits to EPFL CS-433)

We consider a classification problem on linearly separable data. Our dataset had an outlier – a point that is very far from the other datapoints in distance. We trained the linear discriminant analysis (LDA), logistic regression and 1-nearest-neighbour classifiers on this dataset. We tested trained models on a test set that comes from the same distribution as training set, but doesn't have any outlier points. After that we removed the outlier and retrained our models.

After retraining, which classifier will **not change** its decision boundary around the test points.

(A) Logistic regression.

(B) 1-nearest-neighbors classifier.

(C) LDA.

(D) None of them.

▶ Exercise 7: Classifier boundary (credits to EPFL CS-189)

Which of these classifiers could have generated the decision boundary here below
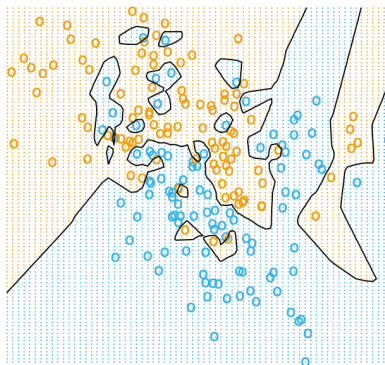
(A) 1-NN

(B) Logistic regression

(C) Quadratic discriminant analysis

(D) None of the above.

▶ Exercise 8 (credits to Berkeley CS-189)

You want to train a dog identifier with Gaussian discriminant analysis. Your classifier takes an image vector as its input and outputs 1 if it thinks it is a dog, and 0 otherwise. You use the CIFAR10 dataset, modified so all the classes that are not "dog" have the label 0. Your training set has 5k dog images and 45k non-dog ("other") images. Which of the following statements seem likely to be correct.

(A) LDA has an advantage over QDA because the two classes have different numbers of training examples.

(B) QDA has an advantage over LDA because the two classes have different numbers of training examples.

(C) LDA has an advantage over QDA because the two classes are expected to have very different covariance matrices.

(D) QDA has an advantage over LDA because the two classes are expected to have very different covariance matrices.

▶ Exercise 9 (credits to CMU 10-701)

For a one-dimensional Gaussian, the probability density looks similar to bell curve. For a two-dimensional Gaussian, if both coordinates are independent of one another then the density concentrates in circles. If the two coordinates are not independent, then the density will look elliptical like. For each dataset below, determine if the Naive Bayes assumption is valid. Assume that the data given the class label is distributed as a multivariate Gaussian.
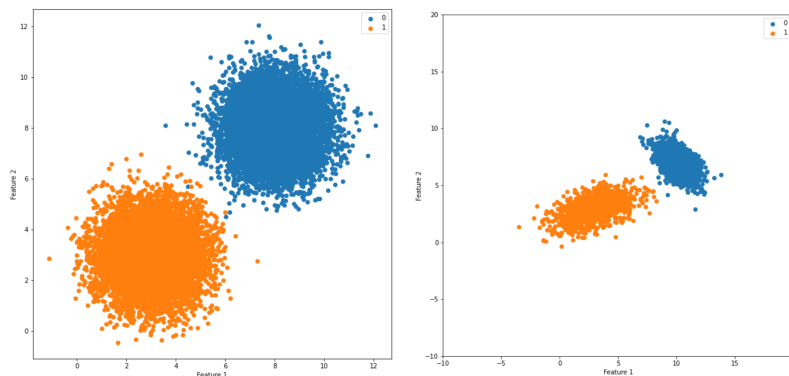
▶ Exercise 10 (credits to CMU 10-701)

Decide whether each of the following sentences is true or false. You must give at least a one sentence explanation for each of your choices.

(01) Decision trees are learned by minimizing information gain (F)

(02) The coefficients α assigned to the classifiers assembled by AdaBoost are always non-negative (F)

(03) Maximizing the likelihood of logistic regression model yields multiple local optima (F)

(04) No classifier can do better than a naive Bayes classifier if the distribution of the data is known (F)

(05) In the discriminative approach to solving classification problems, we model the conditional probability of the labels given the observations (T)

(06) Students from ENSIMAG and PHELMA are trying to solve the same logistic regression problem for a dataset. The PHELMA group claims that their initialization point will lead to a much better optimum than ENSIMAG's initialization point. PHELMA is correct. (F)

(07) In AdaBoost we start with a Gaussian weight distribution over the training samples (F)

(08) A 1-NN classifier has higher variance than a 3-NN classifier (T)