Ensimag – Grenoble INP – UGA
Introduction to Statistical Learning and Applications
Pedro L. C. Rodrigues
Alexandre Wendling
Year 2023-2024
pedro.rodrigues@inria.fr
alexandre.wendling@univ-grenoble-alpes.fr

TD 4: Some questions from previous exams

▶ Exercise 1 (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 2 (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 mean for both classes.

(B) Different covariance matrix for each class.

(C) Same covariance matrix for both classes.

(D) None of the above.

▶ Exercise 3 (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 $\hat{\pi}_0 = 5/6$, then the LDA and QDA classifiers are identical.

(D) None of the above.

▶ Exercise 4 (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 5 (credits to Berkeley CS-189)

How does the bias-variance decomposition of a ridge regression estimator compare with that of ordinary least squares regression?

(A) Ridge has larger bias, larger variance.

(B) Ridge has larger bias, smaller variance.

(C) Ridge has smaller bias, larger variance.

(D) Ridge has smaller bias, smaller variance.

▶ Exercise 6 (credits to EPFL CS-433)

Assume we are doing linear regression with mean-squared loss and L2-regularization on four one-dimensional data points. Our prediction model can be written as $f(x) = ax + b$ and the optimization problem can be written as

$$a^\star, b^\star = \underset{a,b}{\operatorname{argmin}} \sum_{i=1}^{4} \Big(y_i - f(x_i)\Big)^2 + \lambda a^2$$

Assume that our data points (x_i, y_i) are $\{(-2, 1), (-1, 3), (0, 2), (3, 4)\}$. What is the optimal value for the bias, $b^\star$?

(A) Depends on the value of λ.

(B) 3

(C) 2.5

(D) None of the above answers.

▶ Exercise 7 (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.

(1) Decision trees are learned by minimizing information gain

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

(3) Maximizing the likelihood of logistic regression model yields multiple local optimums

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

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

(6) 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.

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

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

(9) If we had infinite data and infinitely fast computers, kNN would be the only algorithm we would study in our course.

(10) The correspondence between logistic regression and Gaussian Naive Bayes (with identity class covariances) means that there is a one-to-one correspondence between the parameters of the two classifiers