Maura Pintor, PhD Student

maurapintor

https://maurapintor.github.io

@maurapintor

https://www.linkedin.com/in/maura-pintor

maura.pintor@unica.it

Exercise 1 (10 points)

Let's consider a 2-class problem in a one-dimensional feature space bounded in [0,1], i.e., $x \in [0,1]$. The class-conditional densities are: $p(x|\omega_1) = 2 - 2x$, and $p(x|\omega_2) = 2x$, both defined in [0,1]. Assume that the prior probability of class 1 is a variable parameter denoted with $P_1 \in [0,1]$.

- (5 points) Compute the optimal Bayesian decision boundary $x^\star$, the Bayesian decision regions and the corresponding classification error E as a function of P_1.
- (3 points) Plot the Bayesian error E as a function of P_1. What kind of dependency (linear, quadratic, cubic) does E show with respect to P_1?
- (2 points) Can you explain why the Bayesian error equals zero under certain conditions?

Exercise 2 (10 points)

Let us consider a 4-class problem in R^2 (two-dimensional feature space), where the likelihood of each class is Gaussian and given as $p(x|\omega_i) = N(\boldsymbol{\mu_i}, \Sigma_i)$, with

$$\Sigma_i = \sigma^2\mathbf{I}\ ;\ \mu_1 = \begin{pmatrix}-1\\-1\end{pmatrix}\ ;\ \mu_2 = \begin{pmatrix}+1\\-1\end{pmatrix};\ \mu_3 = \begin{pmatrix}+1\\+1\end{pmatrix}\ ;\ \mu_4 = \begin{pmatrix}-1\\+1\end{pmatrix}$$

and prior probabilities $P_1 = P_2 = P_3 = P_4$.

Compute the decision boundaries and plot them.

Exercise 3 (5 points)

Given the two-dimensional training points $\mathbf{x}_{\text{tr}}$, along with their labels $\mathbf{y}_{\text{tr}}$, and a set of test examples $\mathbf{x}_{\text{ts}}$, with their labels $\mathbf{y}_{\text{ts}}$: $\mathbf{x}_{\text{tr}} = \begin{bmatrix} -1 & 1 \\ 1 & 1 \\ 1 & -1 \end{bmatrix}$, $\mathbf{y}_{\text{tr}} = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}$, $\mathbf{x}_{\text{ts}} = \begin{bmatrix} -0.5 & 1.5 \\ 0.1 & 0.5 \\ -1 & -1.5 \\ 1 & -0.5 \end{bmatrix}$, $\mathbf{y}_{\text{ts}} = \begin{bmatrix} 2 \\ 1 \\ 1 \\ 0 \end{bmatrix}$, classify the points in $\mathbf{x}_{\text{ts}}$ with a k-NN algorithm with k=1, using the l2 distance as the distance metric. The distance matrix (of the squared values) computed by comparing $\mathbf{x}_{\text{ts}}$ against $\mathbf{x}_{\text{tr}}$ is given below:

```
[[ 0.50   2.50   8.50]
 [ 1.46   1.06   3.06]
 [ 6.25  10.25   4.25]
 [ 6.25   2.25   0.25]]
```

- (3 points) Compute the classification error. In case of equal (minimum) distances between a given test sample and a subset of the training points, assign the test sample to the class of the first point of the training set (from left to right in the distance matrix).
- (2 points) Plot the decision function of this k-NN classifier, along with the training and test data.

Exercise 4 (5 points)

Given the two-dimensional data points $\mathbf{x} = \begin{bmatrix} -2 & 0 \\ 0 & -2 \\ -2 & -1 \\ -1 & 1 \\ 1 & 0 \\ -2 & -1 \end{bmatrix}$ and the initial k=2 centroids $\mathbf{v} = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}$:

- (4 points) Cluster the data points **x** using the k-means algorithm, reporting the clustering labels, the updated centroids, and the objective function at each iteration of the algorithm. Use the L1 (Manhattan) distance for computing the distances between the data points **x** and the centroids **v**. Compute the objective function as $\sum_{i=1}^{n} \left|\left|\boldsymbol{x}_i - \boldsymbol{v}_{\mathbf{k}}\right|\right|_1$ where k is the index of the closest centroid to $\boldsymbol{x}_{\boldsymbol{i}}$. If a point has the same distance with respect to multiple centroids, assign it to the centroid with the lowest class index.
- (1 points) Make a two-dimensional plot displaying the data points and the centroids, indicating to which cluster each point belongs to, after the last iteration.

Links used today

- MachineLearningCheatSheet.pdf
- ML-tutor-07-whiteboard