Maura Pintor, PhD Student

maurapintor

https://maurapintor.github.io

@maurapintor

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

maura.pintor@unica.it

Implementation of the ramp loss exercise:
https://colab.research.google.com/drive/1YCJfWM6ps2d0z7xQr3lJdiry1qG80w7a?usp=sharing

Exercise 2 (12 points)

Let us assume that the probability distributions are defined as in Exercise 1. We want to minimize the classification error (not the risk!) using the rejection option with the Chow's rule.

Let us assume that the rejection region is defined as [-2, 6].

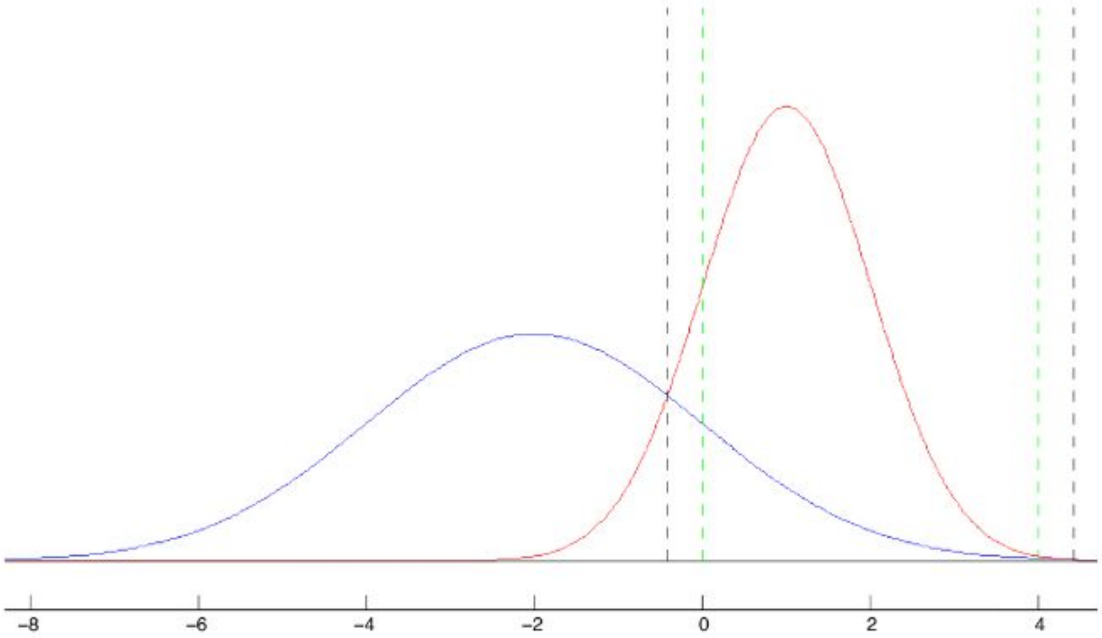
- Compute the rejection threshold T on the posterior probabilities, for class 1 and class 2 on both the bounds of the reject region.
- $\boxed{\text{Why the values of T computed for the two classes are different? Plot (approximately) the posterior distributions w.r.t. x to find the answer.}}$
- Compute the fraction of rejected samples of class 2.

How to plot the posterior distributions?

https://github.com/unica-ml/ml/blob/master/exams/20190415-pre1.pdf

Therefore, the Bayesian decision regions are: $R_1 \equiv (-\infty, x_1) \cup (x_2, +\infty);\ R_2 \equiv [x_1, x_2]$

The plot shows the Bayesian decision boundaries in black, and the minimum-risk ones in green. It is clear from the plot that, when considering the minimum-risk decision regions, the error on class 2 increases, as R_2 is reduced (due to a larger cost of misclassifying class 1 as 2).



Partial solution to the exercise

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 probabilities of the two data classes are $P_1 = P_2$.

- (2 points) Compute the optimal Bayesian decision boundary $x^\star$, the Bayesian decision regions and the corresponding classification error.
- (3 points) Plot the joint distributions $P_k p(x|\omega_k)$ for k=1,2 on the given one-dimensional feature space, along with the Bayesian decision regions. Highlight the Bayesian error area and the additional error area that may be incurred if the decision boundary $x^\star$ is shifted (left or right).
- (5 points) Compute the classification error E as a function of the decision boundary $x^\star$, and plot its behavior against the decision boundary $x^\star \in [0,1]$.
    - What kind of dependency (linear, quadratic, cubic) does E show with respect to $x^\star$?
    - Can you provide an intuitive geometric explanation on why this happens, by looking at the highlighted error areas from the previous plot?

Exercise 2 (10 points)

Let us consider a 2-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

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

and prior probabilities $P_1 = 3P_2$.

Compute the decision boundary and plot it (for a given value of σ^2).

What is the value of σ^2 for which we have the decision boundary corresponding to $x_1 + x_2 = 1$?

Exercise 3 (5 points)

Given the two-dimensional training points $\mathbf{x}_{\mathrm{tr}}$, along with their labels $\mathbf{y}_{\mathrm{tr}}$, and a set of test examples $\mathbf{x}_{\mathrm{ts}}$, with their labels $\mathbf{y}_{\mathrm{ts}}$: $\mathbf{x}_{\mathrm{tr}} = \begin{bmatrix} -1 & 1 \\ 1 & -1 \\ 2 & 1 \\ 1 & 3 \\ 3 & -1 \end{bmatrix}$, $\mathbf{y}_{\mathrm{tr}} = \begin{bmatrix} 2 \\ 2 \\ 1 \\ 1 \\ 0 \end{bmatrix}$, $\mathbf{x}_{\mathrm{ts}} = \begin{bmatrix} 0 & 0 \\ 3 & 2 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}$, $\mathbf{y}_{\mathrm{ts}} = \begin{bmatrix} 2 \\ 1 \\ 2 \\ 0 \end{bmatrix}$, classify the points in $\mathbf{x}_{\mathrm{ts}}$ 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 $\mathbf{x}_{\mathrm{ts}}$ against $\mathbf{x}_{\mathrm{tr}}$ is given below:

```
[[ 2.   2.   5. 10. 10.]
 [17. 13.   2.   5.   9.]
 [ 1.   5.   4.   5. 13.]
 [ 5.   1.   2.   9.   5.]]
```

- (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 $\mathbf{x} = \begin{bmatrix} -2 & 0 \\ 1 & 1 \\ 1 & 2 \\ -1 & 1 \\ -3 & -1 \\ 2 & 1 \end{bmatrix}$ and the initial k=2 centroids $\mathbf{v} = \begin{bmatrix} -1 & 0 \\ 1 & 1 \end{bmatrix}$:

- (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-06-whiteboard]