RPZ Test 7.2.2023

Jméno:

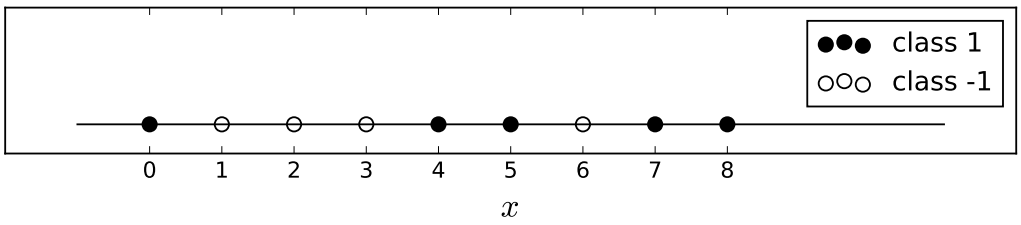
1. (2b) Bayesovská klasifikace. Uvažujte dvě třídy 1 a 2, které mají obě normální rozdělení, $N(\mu_1, \mathbf{C}_1)$ a $N(\mu_2, \mathbf{C}_2)$, a stejné apriorní pravděpodobnosti. Parametry rozdělení jsou:
   $\mu_1 = \begin{bmatrix} 3 \\ 6 \end{bmatrix}$, $\mathbf{C}_1 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$, $\mu_2 = \begin{bmatrix} 3 \\ -2 \end{bmatrix}$, $\mathbf{C}_2 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$.
   Najděte klasifikátor, který minimalizuje chybu klasifikace.

2. (5b) Metoda maximální věrohodnosti (Maximum Likelihood, ML).

   (a) (3b) Životnost žárovek je popsána hustotou pravděpodobnosti $p(t) = \frac{1}{\theta} e^{-t/\theta}, t \in [0, \infty]$, kde $p(t)\delta$ je pravděpodobnost, že žárovka praskne v intervalu $(t, t+\delta)$ $(\delta \to 0)$. Jaký je maximálně věrohodný odhad θ, pokud máme zaznamenán čas selhání šesti žárovek: $t_0 = 0.0, t_1 = 1.0, t_2 = 2.3, t_3 = 2.7, t_4 = 6.0, t_5 = 12.0$?

   (b) (2b) Zařízení zaznamenávající časy selhání žárovek bylo z úsporných důvodů vyřazeno. Žárovky byly napájeny nepřetržitě po dobu T a jedinou informaci, kterou máme k dispozici, je že po této době je G žárovek ještě dobrých (Good) a F žárovek už selhalo (Failed). Najděte maximálně věrohodný odhad θ.

3. (4b) Adaboost. Máme následující jednodimenzionální data:

   a množinu slabých klasifikátorů $h(x) = \text{sign}(ax + b)$ $(a \in \{-1, 1\}, b \in \mathbb{R})$. Vysvětlete na tomto příkladu, jak funguje a co dělá Adaboost (udělejte celou jednu iteraci, končící prvním převálením dat.)

4. (4b) Vysvětlete algoritmus K-means++. Dejte příklad, kde použití K-means++ pomůže.

5. (5b) Rozhodovací stromy.

   (a) (2b) Vysvětlete jednoduchý způsob konstrukce stromu shora-dolů, který používá k výběru rozhodujícího kritéria maximalizaci informačního zisku.

   (b) (3b) Diskutujte výhody a nevýhody rozhodovacích stromů.

RPZ Test Feb 7, 2023

Name:

1. (2 points) Bayes Classification. Consider two classes with equal prior probabilities and Gaussian class-conditional distributions $N(\mu_1, \mathbf{C}_1)$ and $N(\mu_2, \mathbf{C}_2)$, with:

   $\mu_1 = \begin{bmatrix} 3 \\ 6 \end{bmatrix}$, $\mathbf{C}_1 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$, $\mu_2 = \begin{bmatrix} 3 \\ -2 \end{bmatrix}$, $\mathbf{C}_2 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$.

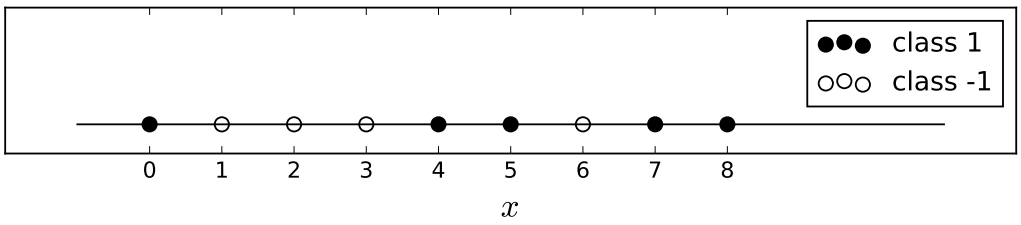
   Find the classifier minimizing the probability of classification error.

2. (5 points) Maximum Likelihood Estimation.

   (a) (3 points) The reliability of light bulbs is defined by the probability density function $p(t) = \frac{1}{\theta} e^{-t/\theta}, t \in [0, \infty]$, where $p(t)\delta$ is the probability that a bulb fails in a short interval δ centered at time t. What is the maximum likelihood estimate of the reliability parameter θ if the following lifetimes have been observed in an experiment with six bulbs: $t_0 = 0.0, t_1 = 1.0, t_2 = 2.3, t_3 = 2.7, t_4 = 6.0, t_5 = 12.0$?

   (b) (2 points) The same problem, but the device for logging the failure times is not available. The bulbs have been powered for a time interval of length T. The only information we have is that after this time, G bulbs are still good and F bulbs have already failed. Find θ.

3. (4 points) Adaboost learning algorithm. Consider the following 1-D data:

   and the following set of weak classifiers: $h(x) = \text{sign}(ax+b)$ ($a \in \{-1, 1\}, b \in \mathbb{R}$). Use this example to explain how Adaboost works (make one full iteration, ending with first data re-weighting.)

4. (4 points) Describe the K-means++ algorithm. Give an example where K-means++ helps.

5. (5 points) Decision tree classifiers.

   (a) (2 points) Describe the simplest top-down tree-building algorithm that selects the decision rule at a node by maximization of information gain.

   (b) (3 points) Discuss advantages and disadvantages of decision trees.