Machine Learning, Spring 2019: Exercise Sheet 3

Problem 1 (easy, informal). Make a list of 5 classification tasks of real-world relevance and present them in a format similar to the table at the beginning of Section 4 of the lecture notes. The purpose of this task is to make you aware of the (almost) universality of the notion of "classification" – once you start thinking of examples you'll find that many relevant real-life problems can be cast as picking the right labels for patterns.

Problem 2. Consider the Digits dataset. Specify (in words or formulas) 8 binary features $f_1, \ldots, f_8$ which assign either the value 1 or the value 0 to a digit image x, such that, if you know $f_1(x), \ldots, f_8(x)$, you know which digit is shown.

Problem 3. Describe a collection of binary decision questions that yield a decision tree for classifying animal classes {**Fish**, **Bird**, **Worm**, **Snake**, **Cat**, **Dog**}.

Problem 4. (computing a decision boundary). Consider a two-class classification problem with classes c_1, c_2, where the decision is based on a single feature f: $\mathbf{P} \to \mathbb{R}$. Let g_1, g_2 be the pdf's of the class-conditional distributions $P_{X \mid Y = ci}$ (where $i = 1, 2$). Concretely, let these pdf's be the Gaussians

$$g_1(x) = \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}, \quad g_2(x) = \frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-\frac{(x-\mu_2)^2}{2\sigma_2^2}}.$$

Let the class probabilities be denoted by

$P(Y = c_i) = \gamma_i \ \ (i = 1, 2)$.

Give a formula for the decision boundary. You will find that there are three distinct cases of how the decision boundary may look like. Draw schematic plots of these three situations.