Exercises for the course

# Machine Learning 1

Winter semester 2020/21

Abteilung Maschinelles Lernen
Institut für Softwaretechnik und theoretische Informatik
Fakultät IV, Technische Universität Berlin
Prof. Dr. Klaus-Robert Müller
Email: klaus-robert.mueller@tu-berlin.de

## Exercise Sheet 1

### Exercise 1: Estimating the Bayes Error $(10 + 10 + 10 \text{ P})$

The Bayes decision rule for the two classes classification problem results in the Bayes error

$$P(\text{error}) = \int P(\text{error}|\boldsymbol{x})\, p(\boldsymbol{x})\, d\boldsymbol{x},$$

where $P(\text{error}|\boldsymbol{x}) = \min\left[\, P(\omega_1|\boldsymbol{x})\,,\; P(\omega_2|\boldsymbol{x})\,\right]$ is the probability of error for a particular input $\boldsymbol{x}$. Interestingly, while class posteriors $P(\omega_1|\boldsymbol{x})$ and $P(\omega_2|\boldsymbol{x})$ can often be expressed analytically and are integrable, the error function has discontinuities that prevent its analytical integration, and therefore, direct computation of the Bayes error.

(a) *Show* that the full error can be upper-bounded as follows:

$$P(\text{error}) \leq \int \frac{2}{\frac{1}{P(\omega_1|\boldsymbol{x})} + \frac{1}{P(\omega_2|\boldsymbol{x})}}\, p(\boldsymbol{x})\, d\boldsymbol{x}.$$

Note that the integrand is now continuous and corresponds to the harmonic mean of class posteriors weighted by $p(\boldsymbol{x})$.

(b) *Show* using this result that for the univariate probability distributions

$$p(x|\omega_1) = \frac{\pi^{-1}}{1 + (x - \mu)^2} \quad \text{and} \quad p(x|\omega_2) = \frac{\pi^{-1}}{1 + (x + \mu)^2},$$

the Bayes error can be upper-bounded by:

$$P(\text{error}) \leq \frac{2\, P(\omega_1)P(\omega_2)}{\sqrt{1 + 4\mu^2 P(\omega_1)P(\omega_2)}}$$

(Hint: you can use the identity $\int \frac{1}{ax^2+bx+c}\, dx = \frac{2\pi}{\sqrt{4ac-b^2}}$ for $b^2 < 4ac$.)

(c) *Explain* how you would estimate the error if there was no upper-bounds that are both tight and analytically integrable. Discuss following two cases: (1) the data is low-dimensional and (2) the data is high-dimensional.

### Exercise 2: Bayes Decision Boundaries $(15 + 15 \text{ P})$

One might speculate that, in some cases, the generated data $p(x|\omega_1)$ and $p(x|\omega_2)$ is of no use to improve the accuracy of a classifier, in which case one should only rely on prior class probabilities $P(\omega_1)$ and $P(\omega_2)$ assumed here to be strictly positive.

For the first part of this exercise, we assume that the data for each class is generated by the univariate Laplacian probability distributions:

$$p(x|\omega_1) = \frac{1}{2\sigma} \exp\Big(-\frac{|x-\mu|}{\sigma}\Big) \quad \text{and} \quad p(x|\omega_2) = \frac{1}{2\sigma} \exp\Big(-\frac{|x+\mu|}{\sigma}\Big).$$

where $\mu, \sigma > 0$.

(a) *Determine* for which values of $P(\omega_1), P(\omega_2), \mu, \sigma$ the optimal decision is to always predict the first class (i.e. under which conditions $P(\text{error}|x) = P(\omega_2|x) \;\; \forall\, x \in \mathbb{R}$).

(b) *Repeat* the exercise for the case where the data for each class is generated by the univariate Gaussian probability distributions:

$$p(x|\omega_1) = \frac{1}{\sigma\sqrt{2\pi}} \exp\Big(-\frac{(x-\mu)^2}{2\sigma^2}\Big) \quad \text{and} \quad p(x|\omega_2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\Big(-\frac{(x+\mu)^2}{2\sigma^2}\Big).$$

where $\sigma > 0$.

### Exercise 3: Programming $(40 \text{ P})$

Download the programming files on ISIS and follow the instructions.