N. Bertschinger
M. Kaschube
V. Ramesh

Machine Learning I
1. Exercise Sheet

Exercise 1. *Show that if two random variables X and Y are independent, then their covariance $\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$ is zero.* **2 points**

Exercise 2. *Suppose that we have three coloured boxes r (red), b (blue), and g (green). Box r contains 3 apples, 4 oranges, and 3 limes, box b contains 1 apple, 1 orange, and no limes, and box g contains 3 apples, 3 oranges, and 4 limes. If a box is chosen at random with probabilities $p(r) = 0.2, p(b) = 0.2, p(g) = 0.6$, and a piece of fruit is removed from the box (with equal probability of selecting any of the items in the box), then what is the probability of selecting an apple? If we observe that the selected fruit is in fact an orange, what is the probability that it came from the green box?* **2 points**

Exercise 3. *Suppose you have observed N samples $x_1, \ldots, x_N$ drawn from a Gaussian distribution. Compute the maximum likelihood estimators for the mean and variance of the data, i.e.*

$$\max_{\mu,\sigma^2} \log \prod_{n=1}^{N} p(x_n; \mu, \sigma^2)$$

where $p(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}$. **2 points**

Exercise 4. *Consider a classification problem with class labels $c_1, \ldots, c_K$. The expected loss of a classification rule $r(x)$ is then given by*

$$\mathbb{E}[l] = \sum_k \int l(r(x), c_k) p(x, c_k) dx$$

Assume that $l(c, c') = \begin{cases} 0 & \text{if } c = c' \\ 1 & \text{otherwise} \end{cases}$. In this case, also called zero-one loss, $\mathbb{E}[l]$ is the missclassification rate. Show that $\mathbb{E}[l]$ is minimized by the rule $r(x) = \texttt{argmax}_{c_k} p(c_k \mid x)$. **2 points**

Exercise 5. *Consider the expected loss for a regression problem under the loss function $l(x, y) = |x - y|$. Show that $\mathbb{E}[l] = \int\int l(y(x), t) p(x, t) dx dt$ is minimized by the conditional median, i.e. the function $y(x)$ such that the probability mass for $t < y(x)$ is the same as for $t \geq y(x)$.* **2 points**