Statistical Machine Learning 2016

Exercises, week 5

29 September 2016

Exercise 1

A factory produces products X. 20% is of quality $x = 1$ and the remainder of quality $x = 2$. There is a test Z, which can have an outcome $\{1, 2, 3, 4, 5\}$. The conditional probability density of z, depending on the quality x is

$p(z = 1|x = 1) = 0.15$; $p(z = 2|x = 1) = 0.15$; $p(z = 3|x = 1) = 0.4$; $p(z = 4|x = 1) = 0.25$; $p(z = 5|x = 1) = 0.05$
$p(z = 1|x = 2) = 0.12$; $p(z = 2|x = 2) = 0.18$; $p(z = 3|x = 2) = 0.2$; $p(z = 4|x = 2) = 0.22$; $p(z = 5|x = 2) = 0.28$

Suppose we observe test result $z = 3$. Compute, using Bayes' rule, the posterior probability $p(x = 1|z = 3)$.

Exercise 2

A factory produces products X. 75% is of quality $x = 1$ and the remainder of quality $x = 2$. There is a test Z, which can be a real number z between 0 and 1. The conditional probability density of z, depending on the quality x is

$$\begin{aligned} p(z|x = 1) &= 2(1 - z) \\ p(z|x = 2) &= 1 \end{aligned}$$

1. Interpret these equations and compute $p(x|z)$ using Bayes' rule
2. Compute the Bayes optimal decision to minimize misclassification rate as function of z, i.e. for which z should one classify $x = 1$ and for which z should one classify $x = 2$.
3. Suppose we have a loss matrix L_{kj}, expressing the loss for classifying as $x = j$ while the true class is k. Suppose this matrix is given by

   $$L_{11} = L_{22} = 0, \quad L_{12} = 1, \quad L_{21} = 5$$

   Compute the optimal decision boundary to minimize expected loss.

Exercise 3

(Bishop 1.22) Given a loss matrix with elements L_{kj}, the expected risk is minimized if, for each $\mathbf{x}$, we choose the class that minimizes:

$$\sum_k L_{kj} p(\mathcal{C}_k|\mathbf{x}) \tag{1}$$

Verify that, when the loss matrix is given by $L_{kj} = 1 - \delta_{kj}$, where δ_{kj} is the Kronecker delta function, this reduces to the criterion of choosing the class having the largest posterior probability. What is the interpretation of this form of loss matrix?

Exercise 4

The Gaussian distribution in one dimension with mean μ and variance σ^2 is

$$\mathcal{N}(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} \tag{2}$$

The Kullback-Leibler divergence $KL(p||q)$ is defined as

$$KL(p(x)||q(x)) = -\int p(x) \ln q(x) dx + \int p(x) \ln p(x) dx \tag{3}$$

Compute the Kullback-Leibler divergence $KL(p||q)$ between two Gaussians with the *same* variance σ^2, but different means μ and m. So $p(x) = \mathcal{N}(x|\mu,\sigma^2)$ and $q(x) = \mathcal{N}(x|m,\sigma^2)$. Verify that $KL(p||q) \geq 0$ and equal if and only if $\mu = m$.

Exercise 5

If a random variable x has distribution $p(x)$, its entropy is

$$H[p(x)] = -\int p(x) \log p(x) dx \tag{4}$$

If two random variables x, y have joint distribution $p(x,y)$, then their entropy is defined as

$$H[p(x,y)] = -\iint p(x,y) \log p(x,y) dx dy \tag{5}$$

Use this to show that:

$$p(x,y) = p(x)p(y) \quad \Rightarrow \quad H[p(x,y)] = H[p(x)] + H[p(y)]$$

Exercise 6

Minimize $f(x,y) = 3x^2 + xy + y^2$ under constraint $x + 2y = 3$.

Exercise 7

For a single binary random variable $x \in \{0,1\}$, with $p(x=1|\mu) = \mu$, the probability distribution over x is known as the Bernoulli distribution

$$p(x|\mu) = \mu^x (1-\mu)^{1-x} \tag{6}$$

1. Show that this distribution satisfies the usual normalization constraint for probabilities, and compute its mean and variance.

For a Bernoulli distributed variable, the loglikelihood function L as function of μ (with $0 \leq \mu \leq 1$) is given by

$$L(\mu) = \ln p(D|\mu) = m \ln \mu + (N-m) \ln(1-\mu) \tag{7}$$

in which $m = \sum_n x_n$.

2. Assuming $0 < m < N$, show that the maximum likelihood solution is given by

$$\mu_{ML} = \frac{m}{N}$$

What do the cases $m = 0$ and $m = N$ represent? Can the solution be extended to cover these as well?

For a discrete, binary random variable x, the entropy is given by

$$H[x] = -\sum_{x\in\{0,1\}} p(x|\mu)\log p(x|\mu) \tag{8}$$

3. Calculate the entropy (in bits) of a throw with a rather bent coin for which $p(\text{heads}) = 2/3$, and compare with a fair coin. $(\log_2(3) \approx 1.6)$

The form of the Bernoulli distribution is not symmetric between the two values of x. Sometimes, it is more convenient to use an equivalent formulation for which $x \in \{-1, 1\}$. The binary distribution over x can then be written in an exponential form

$$p(x|\theta) = \frac{1}{Z(\theta)}\exp(x\theta) \tag{9}$$

with parameter $-\infty < \theta < \infty$.

4. Compute $Z(\theta)$. What is roughly the chance on $x = -1$ when $\theta \approx 1$?