

SML: Exercise 2

Rinor Cakaj, Patrick Nowak



TECHNISCHE
UNIVERSITÄT
DARMSTADT

Summer Term 2020
Sheet 1

Task 1.1: Density Estimation

We are given data C1 and C2, which we suppose to be generated by 2D-Gaussians with parameters μ_1, Σ_1 and μ_2, Σ_2 , respectively.

1.1a)

Assume we are given iid. datapoints $x_i, i = 1, \dots, n$ which are generated by a 2D-Gaussian. Following the max-likelihood principle, we maximize the log-likelihood function

$$l(\mu, \Sigma, x_1, \dots, x_n) = \ln\left(\prod_{i=1}^n p(x_i|\mu, \Sigma)\right) = \sum_{i=1}^n \ln(p(x_i|\mu, \Sigma))$$

for the Gaussian probability density

$$p(x|\mu, \Sigma) = \frac{1}{\sqrt{(2\pi)^k |\Sigma|}} \exp\left(-\frac{1}{2}(x - \mu)^T \Sigma (x - \mu)\right). \quad (1)$$

We receive

$$l(\mu, \Sigma) := l(\mu, \Sigma, x_1, \dots, x_n) = \sum_{i=1}^n \left(-\frac{k}{2} \ln(2\pi) - \frac{1}{2} \ln(|\Sigma|) - \frac{1}{2} (x_i - \mu)^T \Sigma (x_i - \mu) \right) \quad (2)$$

$$= -\frac{nk}{2} \ln(2\pi) - \frac{n}{2} \ln(|\Sigma|) - \frac{1}{2} \sum_{i=1}^n (x_i - \mu)^T \Sigma (x_i - \mu). \quad (3)$$

We compute the derivatives w.r.t. μ and Σ and set them equal to zero. This yields

$$\begin{aligned} \frac{d}{d\mu} l(\mu, \Sigma, x_1, \dots, x_n) &= \frac{d}{d\mu} \left(-\frac{1}{2} \sum_{i=1}^n (x_i - \mu)^T \Sigma (x_i - \mu) \right) \\ &= -\sum_{i=1}^n \frac{d}{d\mu} \frac{1}{2} (x_i - \mu)^T \Sigma (x_i - \mu). \end{aligned}$$

Using the matrix identity $\frac{d}{dw} \frac{w^T A w}{dw} = 2Aw$ which holds if w does not depend on A and if A is symmetric, we get (with $w = (x - \mu)$, $dw = -d\mu$)

$$\begin{aligned} 0 &\stackrel{!}{=} \frac{d}{d\mu} l(\mu, \Sigma, x_1, \dots, x_n) \\ 0 &\stackrel{!}{=} -\sum_{i=1}^n \Sigma^{-1} (x_i - \mu). \end{aligned}$$

Finally, we use that Σ^{-1} is positive definite, so we can leave it out here and get

$$0 \stackrel{!}{=} n\mu - \sum_{i=1}^n x_i ,$$

which is solved for the MLE-estimate

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i . \quad (4)$$

Secondly, we need to compute the derivative w.r.t Σ . To do that, we will need some results from mathematical classes. The following is used without prove:

- Cyclic permutations of a matrix product do not change the trace of it:

$$\text{tr} [ABC] = \text{tr} [CAB]$$

- The trace of a scalar is the scalar itself. In particular: the result of a quadratic form $x^T A x$ is a scalar, such that:

$$x^T A x = \text{tr} [x^T A x] = \text{tr} [x^T x A]$$

- $\frac{d}{dA} \text{tr} [AB] = B^T$
- $\frac{d}{dA} \ln |A| = A^{-T}$