

Machine Learning, Spring 2017: Exercise Sheet 5

Please send your type-set solutions by email to our two TA's Xu He ("Owen") x.he@jacobs-university.de and Felix Schmoll f.schmoll@jacobs-university.de. Join into groups of two or three and submit a single solution sheet per group, indicating the group members' names on the sheet.

Deadline for submission is Friday March 31, 23:59 hrs (email sending timestamp). Submissions arriving later (even a second after midnight) will be corrected but not counted for the course grade.

This problem sheet is a refresher for basic probability concepts. You can easily find solutions for these basic problems on the web, even on Wikipedia, – it's of course a much more profound learning experience when you work out the derivations yourself. But it's up to you...

Problem 1 Give a derivation for the formula $\text{Cov}(X, Y) = E[XY] - E[X]E[Y]$.

$$\begin{aligned}\text{Cov}(X, Y) &= E[(X - E[X])(Y - E[Y])] \\ &= E[XY - XE[Y] - E[X]Y + E[X]E[Y]]\end{aligned}$$

Solution.

$$\begin{aligned}&= E[XY] - E[XE[Y]] - E[E[X]Y] + E[X]E[Y] \\ &= E[XY] - E[X]E[Y] - E[X]E[Y] + E[X]E[Y] \\ &= E[XY] - E[X]E[Y]\end{aligned}$$

Problem 2. Prove that the mean minimizes the quadratic loss, that is, for a random variable X with values in \mathbb{R} ,

$$E[X] = \operatorname{argmin}_{x \in \mathbb{R}} E[(x - X)^2]$$

(this is another good reason for why the quadratic loss is so popular!)

Solution. The expression $E[(x - X)^2]$ is equal to $E[x^2 - 2xX + X^2] = x^2 - E[2xX] + E[X^2] = x^2 - 2xE[X] + E[X^2]$. This is a quadratic polynomial in x . Finding the x argument that minimizes this expression can be done by finding the zero of the first derivative of this expression. The first derivative w.r.t. x is $2x - 2E[X]$. The obvious zero for x is $x = E[X]$.

Problem 3. Show that for two RVs X, Y with values in \mathbb{R} , it holds that $-1 \leq \text{Corr}(X, Y) \leq 1$. (Assuming that both RVs don't have zero standard deviation, and that their joint distribution is characterized by a pdf $f(x, y)$). You may use the following fact (a special case of the so-called *Cauchy-Schwarz* inequality):

$$\left(\int_{\mathbb{R}^2} xy f(x, y) d(x, y) \right)^2 \leq \int_{\mathbb{R}^2} x^2 f(x, y) d(x, y) \cdot \int_{\mathbb{R}^2} y^2 f(x, y) d(x, y),$$

where f is a pdf on \mathbb{R}^2 .

Solution.

By definition, $Corr(X,Y) = \frac{Cov(X,Y)}{\sigma(X)\sigma(Y)} = \frac{E[\bar{X}\bar{Y}]}{\sqrt{E[\bar{X}^2]E[\bar{Y}^2]}}$. In order to show that it ranges

between -1 and $+1$, it is enough to show that its square does not exceed 1, that is, have to show that $\frac{E[\bar{X}\bar{Y}]^2}{E[\bar{X}^2]E[\bar{Y}^2]} \leq 1$, which is equivalent to $E[\bar{X}\bar{Y}]^2 \leq E[\bar{X}^2]E[\bar{Y}^2]$.

Writing these expectations out in their pdf-based integrals gives exactly the Cauchy-Schwarz inequality from the problem statement.