# Machine Learning I — Exercise 3

Sacha Sokoloski, Sebastian Damrich – based on material by Dmitry Kobak 27 Oct 2025.
Deadline: 03 Nov 2025 6pm

**Task 1** Construct a $4 \times 3$ matrix that has rank 2 and one with rank 1. Construct the two matrices so that only one may have zero entries and the other has no repeated entry. Explain why they have the correct rank.

**Task 2** Starting from the likelihood function for linear regression

$$\prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \exp\Big[-\frac{(y_i - \boldsymbol{\beta}^\top \mathbf{x}_i)^2}{2\sigma^2}\Big],$$

derive the maximum likelihood solution for $\sigma^2$
*Hint:* The maximum likelihood solution for $\boldsymbol{\beta}$ was discussed in the lecture.

**Task 3** Consider the task of fitting a Gaussian distribution with mean $\mu$ and variance $\sigma^2$ to the given sample $\{y_i\}_{i=1}^n$ via the maximum likelihood principle. How can one see it as a special case of a regression problem? Apply the general regression formulas to this special case to write down maximum likelihood solutions $\hat{\mu}$ and $\hat{\sigma}^2$.
*Hint:* You should get sample mean and sample variance as solutions for $\hat{\mu}$ and $\hat{\sigma}$.

**Task 4** Let $\boldsymbol{\epsilon}$ be a vector whose elements are independent random variables with zero mean and variance one. What is the expected value of $\mathbf{A}\boldsymbol{\epsilon}$ where $\mathbf{A}$ is some matrix? What is the expected value of of $\|\boldsymbol{\epsilon}\|^2$? Explain your answers.
*Hint:* Your answer must not depend on a specific choice of $\mathbf{A}$, but should hold for any matrix $\mathbf{A}$.

**Task 5*** Assume the following linear generative model

$$y = f(x) + \varepsilon = \mathbf{x}^\top \boldsymbol{\beta} + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2) \tag{1}$$

for some $\boldsymbol{\beta} \in \mathbf{R}^{p+1}$ and features $\mathbf{x} = (1, x, x^2, \ldots, x^p)$. Consider polynomial regression with features $\mathbf{x}' = (1, x, \ldots, x^{p'})$ with $n - 1 \geq p' \geq p$ trained on some train data $D = \{(x^{(i)}, y^{(i)})\}_{i=1}^n$. You may assume that the design matrices of the training data have full rank. Show that the bias at a test point $x$ is zero:

$$\text{bias}(x)^2 = \Big(f(x) - \mathbb{E}_D\big(\hat{f}_D(x)\big)\Big)^2 = 0$$

In addition, explain where your proof goes wrong if $p > p'$.
*Hint:* Use the law of total expectation: $\mathbb{E}_{\mathbf{X},\mathbf{y}}(\cdot) = \mathbb{E}_{\mathbf{X}}(\mathbb{E}_{\mathbf{y}|\mathbf{X}}(\cdot \mid \mathbf{X}))$, where in $\mathbb{E}_{\mathbf{y}|\mathbf{X}}(\cdot)$ you can treat $\mathbf{X}$ as a constant and only consider expectation with respect to $\mathbf{y}$. In this particular case, the expectation with respect to $\mathbf{X}$ will not actually depend on the distribution of $\mathbf{X}$. Note that you can use Task 4 to move terms depending only on $\mathbf{X}$ out of the expectation $\mathbb{E}_{\mathbf{y}|\mathbf{X}}$.