

MACHINE LEARNING I — EXERCISE 2

Sacha Sokoloski, Sebastian Damrich – based on material by Dmitry Kobak

20 Oct 2025.

Deadline: 27 Oct 2025 6pm

Task 1 Let a feature $\mathbf{x}_k \in \mathbb{R}^n$ be orthogonal to the intercept feature $\mathbf{x}_0 = (1, 1, \dots, 1)^\top \in \mathbb{R}^n$. Show that this implies that it has mean zero.

Reminder: Two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^d$ are called *orthogonal* if their inner product (written $\langle \mathbf{u}, \mathbf{v} \rangle$ or $\mathbf{u}^\top \mathbf{v}$) is zero, i.e., $0 = \langle \mathbf{u}, \mathbf{v} \rangle = \sum_{i=1}^d u_i v_i$.

“Showing a statement” means providing a formal proof (e.g. by calculation) that the assumptions imply the conclusion.

Task 2 Let two centered (i.e. having mean zero) features \mathbf{x}_i and \mathbf{x}_j be orthogonal to each other. Show that this implies they have sample Pearson correlation zero.

Reminder: The sample Pearson correlation for n paired data samples $(u_1, v_1), \dots, (u_n, v_n)$ is given by

$$\frac{\sum_{i=1}^n (u_i - \bar{u})(v_i - \bar{v})}{\sqrt{\sum_{i=1}^n (u_i - \bar{u})^2} \sqrt{\sum_{i=1}^n (v_i - \bar{v})^2}}, \quad (1)$$

where \bar{u} and \bar{v} denote means of u_1, \dots, u_n and v_1, \dots, v_n , respectively.

Task 3 Let all features in a linear regression model have mean zero (apart from the intercept feature \mathbf{x}_0). Show that this implies that $\hat{\beta}_0 = \bar{y} = \frac{1}{n} \sum_i y_i$.

Task 4 What is the inverse of a matrix $\mathbf{V}\mathbf{D}\mathbf{V}^\top$, where \mathbf{V} is orthogonal and \mathbf{D} is diagonal with non-zero entries on the diagonal? Explain your answer!

Reminder: A square matrix \mathbf{V} is called *orthogonal* if it has orthonormal columns ($\mathbf{V}^\top \mathbf{V} = \mathbf{I}$). One can show that this implies that it also has orthonormal rows ($\mathbf{V}\mathbf{V}^\top = \mathbf{I}$).

The next exercise is starred (*). Starred exercises are more difficult and optional.

Task 5 (Bishop 3.3)* Consider least squares problem with weights. For each training sample i there is a weight w_i entering the loss function as follows:

$$\mathcal{L} = \frac{1}{n} \sum_{i=1}^n w_i (y^{(i)} - \boldsymbol{\beta}^\top \mathbf{x}^{(i)})^2.$$

Find the explicit formula for $\hat{\boldsymbol{\beta}}$. (Hint: compute the gradient, write it using matrix notation, set to zero. It may be convenient to represent weights as a diagonal square $n \times n$ matrix.)