CS 480/680 Assignment 2
Released Jan 22, 2026
©Marina Meilă
mmp@stat.washington.edu

Problem 1 – Least Squares

Let $x_1, x_2, \dots x_n$ be real numbers, and define by $g(z)$ the function

$$g(z) \;=\; \sum_{i=1}^{n}(x_i - z)^2$$

1. Show that the minimum of g is attained for

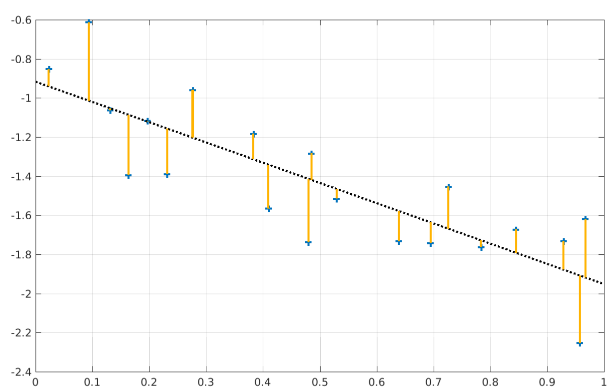
$$z^* \;=\; \frac{1}{n}\sum_{i=1}^{n} x_i$$

[Hint: Take the derivative of g w.r.t. z and solve the equation $g'(z) = 0$.]

2. What is the value $g(z^*)$?

Problem 2 – Linear regression interpretation

No proofs required for this problem



The figure above shows $n = 20$ data points and a linear regression line $f(x) =$

$\beta_0 + \beta_1 x$ estimated by Maximum Likelihood from these data. Answer based on the figure (approximative values to 1 decimal point OK).

1. What are the numerical values of $\beta_{0,1}$?
2. What is the residual for data point with $x^i \approx 0.1$?
3. What is the predicted value $f(x = 0.85)$?
4. True or false: $\sum_{i=1}^{n}(y^i - f(x^i)) = 0$ for any data set? (where f retrained on this data set by Maximum Likelihood)

Problem 3 – Regression Implementation

Recall that ridge regression refers to

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^d} \overbrace{\underbrace{\tfrac{1}{2n}\|X\boldsymbol{\beta} - \mathbf{y}\|_2^2}_{\text{error}} + \lambda\|\boldsymbol{\beta}\|_2^2}^{\text{loss}}, \tag{1}$$

where $X \in \mathbb{R}^{n \times d}$ and $\mathbf{y} \in \mathbb{R}^n$ are the given dataset and $\lambda \geq 0$ is the regularization hyperparameter. If $\lambda = 0$, then this is the standard linear regression problem. Observe the distinction between the error (which does not include the regularization term) and the loss (which does).

1. Show that ridge regression can be rewritten as a non-regularized linear regression problem. That is, prove (1) is equivalent to

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^d} \tfrac{1}{2n} \left\| \begin{bmatrix} X \\ \sqrt{2\lambda n} I_d \end{bmatrix} \boldsymbol{\beta} - \begin{bmatrix} \mathbf{y} \\ \mathbf{0}_d \end{bmatrix} \right\|_2^2, \tag{2}$$

   where I_d is the d-dimensional identity matrix, and $\mathbf{0}_k$ is the zero column vector in k dimensions.
2. Show that the derivative of (1) is

$$\frac{\partial}{\partial \boldsymbol{\beta}} = \tfrac{1}{n} X^\top (X\boldsymbol{\beta} - \mathbf{y}) + 2\lambda\boldsymbol{\beta} \tag{3}$$

3. Implement ridge regression using the closed form solution for linear regression as derived in lecture. You may find the function `numpy.linalg.solve` useful.