Proofs, Exercises and Literature - Regression

1 Proofs

Lemma 1. *The squared L_2-norm $f : \mathbb{R}^d \to \mathbb{R}$, $f(\mathbf{x}) = \|\mathbf{x}\|^2$ is convex.*

Proof. Let $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^d$ and $\alpha \in [0, 1]$. Then we have to show that

$$\|\alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2\|^2 \leq \alpha\|\mathbf{x}_1\|^2 + (1-\alpha)\|\mathbf{x}_2\|^2 \tag{1}$$

We apply the binomial formula for the squared L_2-norm, which derives directly from the definition of the squared L_2-norm by an inner product (see linear algebra lecture). Then we have:

$$\begin{aligned}
\|\alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2\|^2 &= \|\alpha\mathbf{x}_1\|^2 + 2\alpha(1-\alpha)\mathbf{x}_1\mathbf{x}_2 + \|(1-\alpha)\mathbf{x}_2\|^2 \\
&= |\alpha|^2\|\mathbf{x}_1\|^2 + 2\alpha(1-\alpha)\mathbf{x}_1\mathbf{x}_2 + |1-\alpha|^2\|\mathbf{x}_2\|^2 \qquad \text{(homogenity of the norm)} \\
&= \alpha^2\|\mathbf{x}_1\|^2 + 2\alpha(1-\alpha)\mathbf{x}_1\mathbf{x}_2 + (1-\alpha)^2\|\mathbf{x}_2\|^2, \qquad (2)
\end{aligned}$$

where the last equation derives from the fact that the squared absolute value of a real value is equal to the squared real value.

What is standing above, is not yet what we want, and it is difficult to see which step has to be taken next to derive the Inequality (1). Hence, we apply a trick. Instead of showing that Eq. (1) holds as it stands, we show that an equivalent inequality holds:

$$\|\alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2\|^2 - \alpha\|\mathbf{x}_1\|^2 - (1-\alpha)\|\mathbf{x}_2\|^2 \leq 0$$

We substitute now the result of Eq. (2) into the term on the left of the inequality above:

$$\begin{aligned}
&\|\alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2\|^2 - \alpha\|\mathbf{x}_1\|^2 - (1-\alpha)\|\mathbf{x}_2\|^2 \\
&\quad = \alpha^2\|\mathbf{x}_1\|^2 + 2\alpha(1-\alpha)\mathbf{x}_1\mathbf{x}_2 + (1-\alpha)^2\|\mathbf{x}_2\|^2 - \alpha\|\mathbf{x}_1\|^2 - (1-\alpha)\|\mathbf{x}_2\|^2 \\
&\quad = -\alpha(1-\alpha)\|\mathbf{x}_1\|^2 + 2\alpha(1-\alpha)\mathbf{x}_1\mathbf{x}_2 - (1-\alpha)(1-1+\alpha)\|\mathbf{x}_2\|^2 \\
&\quad = -\alpha(1-\alpha)\|\mathbf{x}_1 - \mathbf{x}_2\|^2 \qquad \text{(binomial formula)} \\
&\quad \leq 0
\end{aligned}$$

This concludes what we wanted to show. □

2 Exercises

1. Imagine you have a dataset which consists of three datapoints (of course this is super unrealistic but if we want to go through regression step by step, then we need a really small example). The data is listed in the following table:

D	x_1	y
1	5	2
2	3	5
3	1	3

In this exercise, you are asked to fit a regression function for specified function classes. That is, you will have to create the design matrix and compute the global optimizer(s) β of the regression objective as presented in the lecture. Plot the regression function.

(a) Fit an affine function to the data

$$f(x) = \beta_1 x + \beta_0.$$

Solution: Affine functions are decomposed into an inner product of a basis function and the regression parameter vector $\boldsymbol{\beta}$ by

$$f(x) = \beta_1 x + \beta_0 = \phi_{aff}(x)^\top \boldsymbol{\beta},$$

where the basis function is defined as

$$\phi_{aff}(x) = \begin{pmatrix} 1 \\ x \end{pmatrix}.$$

The design matrix given by

$$X = \begin{pmatrix} - & \phi_{aff}(5)^\top & - \\ - & \phi_{aff}(3)^\top & - \\ - & \phi_{aff}(1)^\top & - \end{pmatrix} = \begin{pmatrix} 1 & 5 \\ 1 & 3 \\ 1 & 1 \end{pmatrix}.$$

We have seen in the lecture videos, that the global minimizer(s) β which minimize the residual sum of squares are given by the solver(s) of the following system of linear equations:

$$\{\boldsymbol{\beta} \mid X^\top X \boldsymbol{\beta} = X^\top y\}. \tag{3}$$

The minimizer β is uniquely defined if $X^\top X$ is invertible. This is here the case.
You can either compute the inverse $(X^\top X)^{-1}$ (via Python or manually) and compute the regression vector β by the formula

$$\boldsymbol{\beta} = (X^\top X)^{-1} X^\top y, \tag{4}$$

or you solve the system of linear equations defined in Eq. (3). We have

$$X^\top X = \begin{pmatrix} 3 & 9 \\ 9 & 35 \end{pmatrix}, \quad X^\top y = \begin{pmatrix} 10 \\ 28 \end{pmatrix}.$$

Hence, we have to solve the following system of linear equations according to Eq. (3):

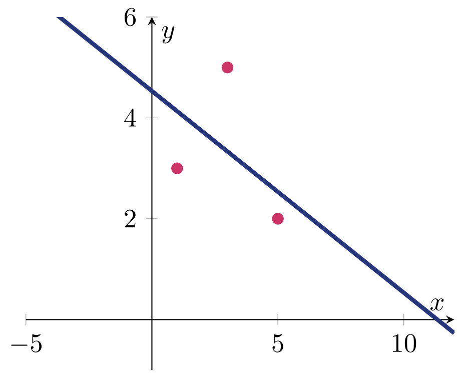
$$\begin{aligned} 3\beta_0 + 9\beta_1 &= 10 \\ 9\beta_0 + 35\beta_1 &= 28. \end{aligned}$$

No matter which way you choose, the result should be

$$\boldsymbol{\beta} \approx \begin{pmatrix} 4.08 \\ -0.25 \end{pmatrix}.$$

Hence, the affine regression function is defined as

$$f(x) = -0.25x + 4.08.$$



(b) Fit a polynomial regression function of degree $k = 2$ to the data

$$f(x) = \beta_2 x^2 + \beta_1 x + \beta_0.$$

Solution: The basis function for a polynomial of degree 2 is here defined as

$$\phi_{p2}(x) = \begin{pmatrix} 1 \\ x \\ x^2 \end{pmatrix}.$$

The design matrix is giiven by

$$X = \begin{pmatrix} - & \phi_{p2}(5)^\top & - \\ - & \phi_{p2}(3)^\top & - \\ - & \phi_{p2}(1)^\top & - \end{pmatrix} = \begin{pmatrix} 1 & 5 & 25 \\ 1 & 3 & 9 \\ 1 & 1 & 1 \end{pmatrix}.$$

We calculate the matrices

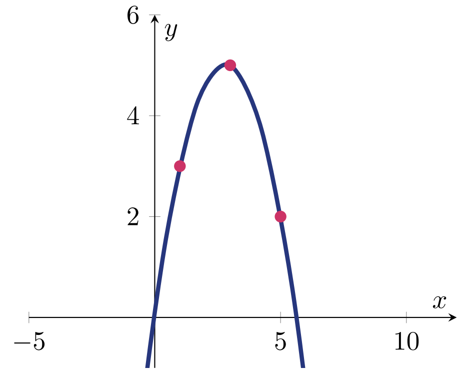
$$X^\top X = \begin{pmatrix} 3 & 9 & 35 \\ 9 & 35 & 153 \\ 35 & 153 & 707 \end{pmatrix}, \quad X^\top y = \begin{pmatrix} 10 \\ 28 \\ 98 \end{pmatrix},$$

and solve the system of linear equation in Eq. (3) or compute the inverse of $X^\top X$ and compute β by Eq. (4). As a result we get

$$\boldsymbol{\beta} \approx \begin{pmatrix} 0.125 \\ 3.5 \\ -0.625 \end{pmatrix}.$$

Hence, the polynomial regression function is

$$f(x) = -0.625x^2 + 3.5x + 0.125.$$



(c) Fit a sum of three Gaussians to the data:

$$f(x) = \beta_1 \exp(-(x-5)^2) + \beta_2 \exp(-(x-3)^2) + \beta_3 \exp(-(x-1)^2).$$

The mean values μ, which have to be specified when we choose a Gaussian basis function, are here equal to the three given feature values in the data. This strategy is also often used in practice.

Solution: The basis function for the Gaussian basis function is defined as

$$\phi_{G3}(x) = \begin{pmatrix} \exp(-(x-5)^2) \\ \exp(-(x-3)^2) \\ \exp(-(x-1)^2) \end{pmatrix}.$$

The design matrix is defined as

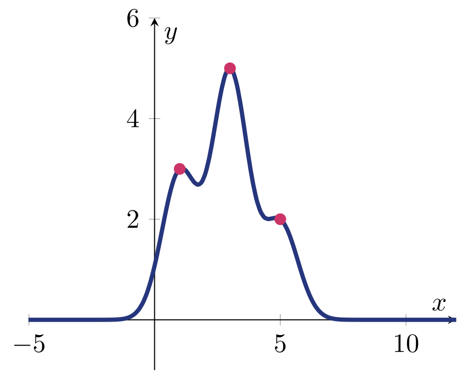
$$X = \begin{pmatrix} - & \phi_{G3}(5)^\top & - \\ - & \phi_{G3}(3)^\top & - \\ - & \phi_{G3}(1)^\top & - \end{pmatrix} = \begin{pmatrix} 1 & \exp(-4) & \exp(-16) \\ \exp(-4) & 1 & \exp(-4) \\ \exp(-16) & \exp(-4) & 1 \end{pmatrix}.$$

I would recommend to use Python to compute the parameter vector according to the formula in Eq. (4), which is applicable here. As a result we get

$$\boldsymbol{\beta} \approx \begin{pmatrix} 1.91 \\ 4.91 \\ 2.91 \end{pmatrix}.$$

Hence, the polynomial regression function is

$$f(x) = 1.91 \exp(-(x-5)^2) + 4.91 \exp(-(x-3)^2) + 2.91 \exp(-(x-1)^2).$$



3 Recommended Literature

As always, the best exercise is to go through the lecture and see if you can follow the steps (maybe with pen and paper). If you feel like reading, I can recommend the following two chapters from the recommended books:

Bishop. Pattern recognition and machine learning. 2006.

3.1 Linear Basis Function Models

- **3.1.1** Maximum Likelihood and Least Squares

3.2 Bias-Variance Decomposition

Friedman, Hastie, and Tibshirani. The elements of statistical learning. 2001.

2.6 Statistical Models, Supervised Learning and Function Approximations

- **2.6.1** A Statistical Model for the Joint Distribution $Pr(X, Y)$
- **2.6.2** Supervised Learning
- **2.6.3** Function Approximation

2.9 Model Selection and Bias-Variance Tradeoff

3 Linear Methods for Regression

- **3.1** Introduction
- **3.2** Linear Regression Models and Least Squares