Data Mining and Machine Learning

Introduction to the basics of linear regression

Gergely Horváth

September 29, 2021

The problem statement – I.

The "reality":

$$y = \sum_{i}^{n} (p_i \cdot x_i) + \epsilon$$

The model:

$$\hat{y} = \sum_{i}^{n} (p_i \cdot x_i)$$

Let us have matrix notations:

$$Y = \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ . \\ . \\ y^{(m)} \end{bmatrix} \qquad X = \begin{bmatrix} x^{(1)} \\ x^{(2)} \\ . \\ . \\ x^{(m)} \end{bmatrix}$$

Where $Y \in \mathbb{R}^{m \times 1}$, $X \in \mathbb{R}^{m \times n}$ and $p \in \mathbb{R}^{n \times 1}$.

The problem statement – II.

Notation repetition:

$$Y = X \cdot p + \epsilon$$
$$\hat{Y} = X \cdot p$$

Next step is to define the error made by the model (in this case the – arguably – simplest convex function will be used):

$$L(D,\ p) = \left\| Y - \hat{Y} \right\|_2 = \| Y - (X \cdot p) \|_2$$

The error is expressed with a 2-norm, so the next step is to minimize the loss, to find the global minimum of this function. → This is going to be easy and trivial, since we have a convex function.

Loss function, objective function, reward function

ALERT!

These functions are created, so we can satisfy some conditions, mathematically speaking. Their roles:

- **Loss function:** generally, express error with their help, the goal is to minimize error
- **Objective function:** a more complex function, that is constructed to express some more abstract optimization goal, we can have e.g. minimization and maximization here as well
- **Reward function:** mainly used in reinforcement learning, an indirect way to give feedback

Examples

- **Loss function:** MSE, MAE or Log-loss (binary cross-entropy)
- **Objective function:** loss functions are a kind of objectie functions, or quadratic cost-function for LQR, and also the regularized ones Ridge or Lasso regressions
- **Reward function:** K-armed bandit, or of a patient's health improvement (w.r.t costs)

Minimization of a convex, 2-norm loss-function

We seek to minimize the loss with respect to our parameters! Elementary calculus: the minimum can be found, where the first derivative is zero. Therefore we need to take the derivative of $L(D, p)$ with respect to p. The following equation must be satisfied:

$$\frac{\partial L(D,\ p)}{\partial p} = 0$$
$$(Y - X \cdot p)^T \cdot X^T = 0$$

Question: the 2-norm has a power of 2. Why can you not see that 2 in the equation?

$$X^T \cdot Y = X^T \cdot X \cdot p$$
$$\cancel{p = (X^T \cdot X)^{-1} \cdot X^T \cdot Y}$$
$$\hat{p} = (X^T X)^{-1} \cdot X^T Y$$