Assignment for Lecture 5

FIRST LESSON FROM LEARNING FROM DATA, CURVE FITTING, AND NO-FREE-LUNCH THEOREM

Lecture Date: 4/1/2026

"C" denotes for "computational" problems, language suggestion: Python/Julia

please include codes and results with analyses for computational problems

please write in pdf format and submit to `bjcai@fudan.edu.cn` ***before the lecture of 4/8/2026***

1. Derive the expressions for w_j's in the case of $n = 2$ and $n = 3$, write down the form of $\mathbf{F}^{-1}$.

2. Write down the equation for $\mathbf{w}^*$ if the regularization term is taken as $\lambda(\mathbf{w}^2)^2$, where $(\mathbf{w}^2)^2 = (w_0^2 + w_1^2 + \cdots + w_n^2)^2$. What's about $\mathbf{w}^4 = w_0^4 + w_1^4 + \cdots + w_n^4$? Moreover, what's about if the regularization term is $\lambda \mathbf{w}^\top \mathbf{T} \mathbf{w}$ with $\mathbf{T}$ a positive-definite matrix?

3. [C] In the linear learning model, define the distance of the point $(x^{(i)}, y^{(i)})$ to the fitting line $y = ax + b$ as,

$$d_\perp^{(i)}(a,b) = \left|a x^{(i)} + b - y^{(i)}\right| \Big/ \sqrt{a^2+1}\,, \quad d_\mathrm{x}^{(i)}(a,b) = \left|a x^{(i)} + b - y^{(i)}\right| \Big/ a. \tag{1-1}$$

The loss function is obtained as $J_\perp(a,b) = 2^{-1}\sum_{i=1}^m d_\perp^{(i),2}(a,b)$ or $J_\mathrm{x}(a,b) = 2^{-1}\sum_{i=1}^m d_\mathrm{x}^{(i),2}(a,b)$. Derive the equations for determining the parameters a and b. See FIG. 1 for the geometrical meaning of $d_\perp$ or d_x, and if a is large $d_\perp \approx d_\mathrm{x}$. Adopt the gradient descent algorithm to find the optimal values for the learning parameters of the loss function (1-1), set the simulation data as given in the lecture notes. Are the predictions for a and b biased or not?

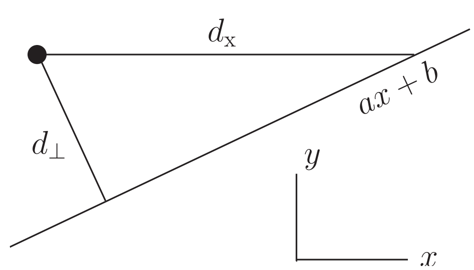

FIG. 1: Distances $d_\perp$ and d_x.