Machine Learning (2022-2023-2)

Homework 1

Problem 1: Let$\{(X_i, y_i^*)\}_{i=1,\ldots,n}$ denote a set of training samples, where $y_i^* \in \{0, 1\}$ is the ground-truth label, and let p_i denote the probability of $Pr(y_i = 1|X_i) = p_i$, which is estimated by the logistic regression model. The likelihood of labels of all samples being correctly estimated is given as

$$\Pr(\beta) = \prod_{i=1}^{n} p_i^{y^*} (1 - p_i)^{1-y_i^*} = \prod_{i=1}^{n} \frac{e^{y_i^* X_i^\intercal \beta}}{1 + e^{X_i^\intercal \beta}}. \tag{1}$$

To maximize the likelihood function $\Pr(\beta)$, we could use gradient ascent on β, *i.e.*,

$$\beta^{\text{new}} = \beta^{\text{old}} + \eta \frac{\partial \log \Pr(\beta)}{\partial \beta},$$

where η denotes the learning rate. Please write the explicit formula for the iterative gradient ascent algorithm. Please explain why the gradient should be computed on $\log \Pr(\beta)$, not on $\Pr(\beta)$.