AIC-4101C – Machine learning

2. Logistic regression

Kevin Zagalo & Marwan W. El Khazen

<kevin.zagalo@inria.fr>

<marwan.wehaiba-el-khazen@inria.fr>

Fall 2021

Exercise 1

In this exercise, $X \in \mathbf{R}^{n\times d}$, $\mathbf{y} \in \{-1, 1\}^n$. A logistic regression consist in finding the optimal parameter $\hat{\mathbf{w}}$, such that $\Pr(y_i = 1|\mathbf{x}_i) = \sigma(\mathbf{x}_i \cdot \hat{\mathbf{w}})$, with $\sigma(z) = (1 + e^{-z})^{-1}$. The log-likelihood for this classification problem is given by

$$\mathcal{L}(\mathbf{w}) = \log \prod_{i=1}^{n} \Pr(y_i = 1|\mathbf{x}_i) = \log \prod_{i=1}^{n} \sigma(\mathbf{x}_i \cdot \mathbf{w})$$

The maximum likelihood estimator is

$$\hat{\mathbf{w}} = \mathrm{argmax}_{\mathbf{w}\in\mathbf{R}^{d+1}} \mathcal{L}(\mathbf{w})$$

Question 1

Show that $\mathcal{L}(\mathbf{w}) = -\frac{1}{n}\sum_{i=1}^{n} \log\big(1 + e^{-y_i(\mathbf{x}_i\cdot\mathbf{w})}\big)$

Question 2

What would be the log-likelihood function for $y_i \in \{0, 1\}$?

Question 3

Let $\ell(y, y') = \log(1 + e^{-yy'})$. From the definition of the likelihood, show that computing the maximum likelihood estimator $\hat{w}$ is equivalent to the optimization problem $\hat{\mathbf{w}} \in \operatorname{argmin}_{\mathbf{w} \in \mathbf{R}^{d+1}} \frac{1}{n} \sum_{i=1}^{n} \ell(y_i, \mathbf{x}_i \cdot \mathbf{w})$ where ℓ is the logistic loss function.

Question 4

Show that the gradient of the logistic loss function defined in log is given by

$$\frac{1}{n} \sum_{i=1}^{n} (\sigma(\mathbf{x}_i \cdot \mathbf{w}) - y_i)\mathbf{x}_i$$

$\blacksquare$