DM n°2 de Machine Learning

Asymptotic variance of logistic regression

The logistic model assumes that the random variables $(X, Y) \in \mathbb{R}^p \times 0, 1$ are such that

$$\mathbb{P}(Y = 1|X) = \frac{\exp(\langle \beta^\star, X \rangle)}{1 + \exp(\langle \beta^\star, X \rangle)} \,,$$

with $\beta^\star \in \mathbb{R}^p$. In this case, the logistic regression estimator $\hat{\beta}_n \in \mathbb{R}^p$ is defined as the Maximum Likelihood Estimator

$$\hat{\beta}_n \in \arg\max_{\beta \in \mathbb{R}^p} \prod_{i=1}^{n} \Big[\big(\frac{\exp(\langle \beta, x_i \rangle)}{1 + \exp(\langle \beta, x_i \rangle)} \big)^{Y_i} \big(\frac{1}{1 + \exp(\langle \beta, x_i \rangle)} \big)^{1 - Y_i} \Big] \,,$$

and defines the classifier

$$\hat{h}_n \; : \; x \mapsto \mathbb{1}_{\langle \hat{\beta}_n, x \rangle > 0} \,.$$

1. Compute the gradient and the Hessian H_n of the negative log-likelihood

$$\ell_n \; : \; \beta \mapsto -\frac{1}{n} \sum_{i=1}^{n} \big[Y_i \langle x_i, \beta \rangle - \log(1 + \exp(\langle x_i, \beta \rangle)) \big] \,.$$

2. What can be said about the function ℓ_n when for all $\beta \in \mathbb{R}^p$, $H_n(\beta)$ is nonsingular?

This assumption is assumed to hold in the following questions.

3. Prove the there exists $\tilde{\beta}_n \in \mathbb{R}^p$ such that

$$\|\tilde{\beta}_n - \beta^\star\|_2 \leq \|\hat{\beta}_n - \beta^\star\|_2 \quad \text{and} \quad \hat{\beta}_n - \beta^\star = H_n(\tilde{\beta}_n)^{-1} \nabla \ell_n(\beta^\star) \,.$$

In the following we assume that $\hat{\beta}_n \to \beta^\star$ almost surely, and the exists a continuous and nonsingular function H such that $H_n(\beta)$ converges to $H(\beta)$, uniformly in a ball around $\beta^\star$.

4. Define for all $i \in [n]$, $p_i(\beta) = \exp(\langle x_i, \beta \rangle)/(1 + \exp(\langle x_i, \beta \rangle))$ and check that for all $t \in \mathbb{R}^p$,

$$\begin{aligned} \mathbb{E}\Big[\exp\big(-\sqrt{n} \langle t, \nabla \ell_n(\beta^\star) \rangle \big) \Big] &= \prod_{i=1}^{n} \Big(1 - p_i(\beta^\star) + p_i(\beta^\star) \exp(\langle t, x_i \rangle / \sqrt{n}) \Big) \exp(-p_i(\beta^\star) \langle t, x_i \rangle / \sqrt{n}) \\ &= \exp\Big(\frac{1}{2} \langle t, H_n(\beta^\star) t \rangle + \varepsilon_n \Big) \,, \end{aligned}$$

where ε_n is a random variable such that for all $\delta > 0$, there exists M, N such that

$$\mathbb{P}(|\sqrt{n} \varepsilon_n| > M) < \delta \,, \quad \forall n > N \,.$$

This is denoted by $\varepsilon_n = \mathcal{O}_{\mathbb{P}}(n^{-\frac{1}{2}})$.

5. Prove that

$$\forall t \in \mathbb{R}^p, \quad \mathbb{E}\Big[\exp\big(-\sqrt{n} \langle t, \nabla \ell_n(\beta^\star) \rangle \big) \Big] \to_{n \to +\infty} \exp\Big(\frac{1}{2} \langle t, H(\beta^\star) t \rangle \Big) \,.$$

It shows that $\sqrt{n} \nabla \ell_n(\beta^\star)$ converges in distribution to $\mathcal{N}_p(0, H(\beta^\star))$. Using Question 3 and Slutsky's lemma one can prove that $\sqrt{n}(\hat{\beta}_n - \beta^\star)$ converges in distribution to $\mathcal{N}_p(0, H(\beta^\star)^{-1})$, this is the asymptotic normality of the logistic regression estimator.