Take home exam 2023

April 21, 2023

Personal work, to be returned on May 12th. Estimated work time: 4-6 hours.

Logistic regression

Let $x \in \mathbb{R}^n$ denote a vector of feature variables, and $y \in \{-1, +1\}$ denote an associated binary outcome. We aim at predicting the outcome y with the help of the feature variables x with the help of a logistic model. For a given parameter $\theta \in \mathbb{R}^n$, the logistic model is defined by

$$P_\theta(y|x) = \frac{1}{1 + \exp(-y \cdot \theta^\top x)} \ ,$$

where $P_\theta(y|x)$ is the conditional probability of y given $x \in \mathbb{R}^n$. Suppose we are given a set of m observations

$$(x_i, y_i) \in \mathbb{R}^n \times \{-1, +1\} \ , \quad \forall i = 1, \cdots, m \ .$$

The likelihood associated with the observations is $\prod_{i=1}^m P_\theta(y_i|x_i)$. We aim at finding the parameter $\theta \in \mathbb{R}^n$ that maximizes the (log)-likelihood.

Formulation of the logistic regression problem

1. Show that maximizing the log-likelihood is equivalent to minimizing the *empirical logistic loss*

$$\min_{\theta \in \mathbb{R}^n} \ L(\theta) \quad \text{with} \quad L(\theta) := \frac{1}{m} \sum_{i=1}^m \log\big(1 + \exp(-y_i \cdot \theta^\top x_i)\big) \ . \tag{1}$$

Problem (1) is called the *logistic regression problem.* We define the *logistic loss function* as $f(z) = \log\big(1 + \exp(-z)\big)$, such that $L(\theta) = \frac{1}{m}\sum_{i=1}^m f(y_i \cdot \theta^\top x_i)$.

2. Show that the logistic regression problem (1) is convex. Is it strongly convex?
3. For a fixed $\theta \in \mathbb{R}^n$, compute the gradient $\nabla_\theta L(\theta)$ and the Hessian $\nabla^2_{\theta\theta} L(\theta)$ of the function L.
4. Discuss which numerical algorithm(s) would be appropriate to solve (1).
5. Show that the conjugate of the logistic loss function f is, for $p \in \mathbb{R}$,

$$f^\star(p) = \sup_{z \in \mathbb{R}} \ \big\{p^\top z - f(z)\big\} = \begin{cases} -p\log(-p) + (1+p)\log(1+p) \ , & \text{if} \quad p \in]-1, 0[\\ 0 \ , & \text{if} \quad p \in \{-1, 0\} \\ +\infty \ , & \text{otherwise} \end{cases} \tag{2}$$

6. Deduce the expression of the conjugate function $L^\star(q)$ of the function L for $q \in \mathbb{R}^n$.

ℓ_1-regularized logistic regression

We add a ℓ_1 regularization to the original logistic regression problem (1). Let $\mu > 0$ be a given parameter. From now on, we consider the problem:

$$\min_{\theta \in \mathbb{R}^n} \; L(\theta) + \mu\|\theta\|_1 = \frac{1}{m}\sum_{i=1}^{m} \log\left(1 + \exp(-y_i \cdot \theta^\top x_i)\right) + \mu \sum_{i=1}^{m} |\theta_i| \; . \tag{3}$$

7. Is the regularized problem (3) convex? Differentiable?
8. Show that an optimal solution $\theta^\sharp$ of (3) satisfies the inclusion

$$0 \in \nabla L(\theta^\sharp) + \mu \partial \|\theta^\sharp\|_1 \; .$$

9. Show that the regularized problem (3) is equivalent to the smooth reformulation

$$\min_{\theta \in \mathbb{R}^n, \alpha \in \mathbb{R}^n} \; L(\theta) + \mu 1^\top \alpha \quad \text{subject to} \quad -\alpha_j \le \theta_j \le \alpha_j \; , \quad \forall j = 1, \cdots, n \; . \tag{4}$$

   Name one algorithm able to solve the constrained problem (4).

Alternatively, using variable lifting the regularized problem is equivalent to

$$\min_{\theta \in \mathbb{R}^n, z \in \mathbb{R}^m} \; \frac{1}{m}\sum_{i=1}^{m} f(z_i) + \mu\|\theta\|_1 \quad \text{subject to} \quad z_i = y_i \cdot x_i^\top \theta \; , \quad \forall i = 1, \cdots, m \; . \tag{5}$$

Let $A \in \mathbb{R}^{m \times n}$ the matrix defined as $A = \begin{bmatrix} y_1 \cdot x_1 \\ \vdots \\ y_m \cdot x_m \end{bmatrix}$.

10. Show that the dual problem of (5) is given by (6). Is there a duality gap?

$$\max_{p \in \mathbb{R}^m} -\frac{1}{m}\sum_{i=1}^{m} f^\star(-m\, p_i) \quad \text{subject to} \quad \|A^\top p\|_\infty \le \mu \; . \tag{6}$$