Prof. Dr. Sören Bartels, Tatjana Schreiber April 30th, 2025

Algorithmic Aspects of Data Analytics and Machine Learning

SS 2025 — Sheet 2

`https://aam.uni-freiburg.de/agba/lehre/ss25/algml/index.html`

Due: May 9, 2025, 2 p.m.

Task 1 (5 points)

We consider a binary classification task using logistic regression. The dataset consists of only two points in $\mathbb{R}$:

$$D = \{(x_1 = -1, y_1 = 0), (x_2 = 1, y_2 = 1)\},$$

where $x_i \in \mathbb{R}$ are the feature values and $y_i \in \{0, 1\}$ are the class labels.

(i) Determine the likelihood function $L(w, b)$ corresponding to the logistic model

$$f_w(x) = \frac{1}{1 + e^{-(wx+b)}}.$$

Calculate $L(-5, 0)$, $L(10, 10)$ and $L(100, 0)$.

(ii) Are the data points linearly separable? Discuss what this implies in terms of the existence of a maximizer for the likelihood function.

(iii) Prove that for every $(w, b) \in \mathbb{R}^2$ the likelihood satisfies $0 < L(w, b) < 1$ and that

$$\sup_{(w,b)\in\mathbb{R}^2} L(w, b) = 1, \qquad \inf_{(w,b)\in\mathbb{R}^2} L(w, b) = 0.$$

Conclude that $L(w, b)$ attains no global maximum at any finite parameter vector. Make a sketch that illustrates this statement.

Task 1 (3 points)

Let $f : \mathbb{R} \to \mathbb{R}$ be defined by

$$f(x) = 3x^4 + 4x^3 - 36x^2 + 42.$$

Demonstrate by performing the gradient descent method with a constant step size that, for an unfavorably chosen starting point, the method converges to a local minimum which is not a global minimizer.

Task 2 (3 points)

Let $f : \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = x^2$. Investigate the convergence of the sequence $(x^{(k)})_{k\in\mathbb{N}}$ defined by

$$x^{(k+1)} = x^{(k)} - 2^{-k} f'(x^{(k)}).$$

Comment on your results.

Task 3 (5 points)

Let $D = \{(x_i, y_i) \in \mathbb{R}^d \times \{0, 1\} \mid i = 1, \dots, n\}$ be a dataset and let ℓ be the associated negative log-likelihood:

$$\begin{aligned}\ell(w) :&= -\log\left(\prod_{i=1}^{n} \sigma(\langle w, x_i\rangle)^{y_i} \left(1 - \sigma(\langle w, x_i\rangle)\right)^{1-y_i}\right)\\ &= \sum_{i=1}^{n}\left(-y_i\langle w, x_i\rangle + \log\left(1 + e^{\langle w, x_i\rangle}\right)\right).\end{aligned}$$

where $\sigma(z) = \frac{1}{1+e^{-z}}$ denotes the sigmoid function.

(i) Show that the gradient of ℓ with respect to w is given by

$$\nabla\ell(w) = \sum_{i=1}^{n} \left(\sigma(\langle w, x_i\rangle) - y_i\right) x_i.$$

(ii) Show further that for the dataset

$$D = \{(-1, 1),\ (0, 0),\ (1, 1)\},$$

the gradient $\nabla\ell$ has exactly one zero, and compute this zero explicitly.

Practical exercise

The following exercise is not mandatory; the points are bonus points that you can collect. Please submit your solutions as a MATLAB or Python file by May 9, 2 p.m., via email to tatjana.stiefken@mathematik.uni-freiburg.de. Please comment your code and your results.

Project (4* points)

Okun's Law is an empirical relationship that describes a negative correlation between changes in the unemployment rate and the growth rate of an economy's output, typically measured as real GDP. The basic idea is that when an economy grows faster than its potential output, the unemployment rate tends to decrease, and vice versa.
A common linear version of Okun's Law is expressed as:

$$\Delta u_t = \beta_0 + \beta_1 \cdot \Delta y_t + \varepsilon_t,$$

where:

- Δu_t is the change in the unemployment rate at time t (i.e., $\Delta u_t = u_{t+1} - u_t$),
- Δy_t is the GDP growth rate at time t (i.e., $\Delta y_t = y_{t+1} - y_t$),
- ε_t is the error term.

Examine Okun's Law using the data provided on the website. Use MATLAB or Python to estimate the coefficients β_0 and β_1 by solving the regression problem. Plot the data points and the resulting regression line.