Exercises Set 6

Paul Dubois

September 29, 2023

Abstract

As this is the last session, there will be no compulsory questions this time.

1 Lagrangian multiplier technique

1.1 Unconstrained optimization

Let $f(x, y) = 2x^2 - 12x + 4y^2 + 8y + 20$.
Find $(x^*, y^*) \in \mathbb{R}^2$ such that f reaches its minimum (i.e. $f(x^*, y^*) \leq f(x, y) \quad \forall (x, y) \in \mathbb{R}^2$).

1.2 (Equality) Constrained optimization

Let $f(x, y) = 2x^2 - 12x + 4y^2 + 8y + 20$.
Suppose further that we want $3x + 5y = 2$.
Find $(x^*, y^*) \in \mathbb{R}^2$ such that $3x^* + 5y^* = 2$ and f reaches its minimum (i.e. $f(x^*, y^*) \leq f(x, y) \quad \forall (x, y) \in \mathbb{R}^2,\ 3x + 5y = 2$).

1.3 Lagrange multiplier

Let $f(x, y) = 2x^2 - 12x + 4y^2 + 8y + 20$.
Suppose further that we want $3x + 5y = 2$.
Let $\mathcal{L}(x, y, \lambda) = f(x, y) - \lambda(3x + 5y - 2)$.
Find the point where $\nabla \cdot \mathcal{L} = 0$

1.4 (Inequality) Constrained optimization

Let $f(x) = x^2 - 2x$.
Suppose further that we want $3x \leq 2$.
Find $x^* \in \mathbb{R}$ such that $3x^* \leq 2$ and f reaches its minimum (i.e. $f(x^*) \leq f(x) \quad \forall x \in \mathbb{R},\ 3x \leq 2$).

Let $f(x) = x^2 + 2x$.
Suppose further that we want $3x \leq 2$.
Find $x^* \in \mathbb{R}$ such that $3x^* \leq 2$ and f reaches its minimum (i.e. $f(x^*) \leq f(x) \quad \forall x \in \mathbb{R},\ 3x \leq 2$).

1.5 Lagrange multiplier

Let $f(x) = x^2 - 2x$.
Suppose further that we want $3x \leq 2$.
Let $\mathcal{L}(x, \lambda) = f(x) - \lambda(3x - 2)$.
Find the point where $\nabla \cdot \mathcal{L} = 0$ and $\lambda \geq 0$.

Let $f(x) = x^2 + 2x$.
Suppose further that we want $3x \leq 2$.
Let $\mathcal{L}(x, \lambda) = f(x) - \lambda(3x - 2)$.
Find the point where $\nabla \cdot \mathcal{L} = 0$ and $\lambda \geq 0$.

2 Support Vector Machines

2.1 Theory

Define a line in $\mathbb{R}^2$ with parameters $\mathbf{w}$ and b defined by $\mathbf{w}.\mathbf{x} = b$ (or $\mathbf{w}.\mathbf{x} - b = 0$) for $\mathbf{x} \in \mathbb{R}^2$. This line cut the plane in 2 regions:

- $\mathbf{x} \in \mathbb{R}^2$ such that $\mathbf{w}.\mathbf{x} - b < 0$
- $\mathbf{x} \in \mathbb{R}^2$ such that $\mathbf{w}.\mathbf{x} - b > 0$

The goal is to find $\mathbf{w}$ and b such that all points of the first class are in the first region, and all points of the second class are in the second region.

2.2 Practice

The training dataset consists of the following data points:
Positive class ("+1"):

- (2, 2)
- (1, 1)

Negative class ("-1"):

- (0, 1)
- (1, 0)

The SVM model's objective function for a linearly separable dataset is as follows:

$$\text{Minimize } \frac{1}{2}\|\mathbf{w}\|^2$$

Subject to the constraints:

$$y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 \text{ for all } i$$

Where:

y_i is the class label $(+1 \text{ or } -1)$ of the i-th data point.
$\mathbf{w}$ is the weight vector normal to the hyperplane.
$\mathbf{x}_i$ is the i-th data point.
b is the bias term.

1. Calculate the optimal values of $\mathbf{w}$ and b to separate the data points while maximizing the margin.
2. Determine the equation of the optimal hyperplane in the form $\mathbf{w} \cdot \mathbf{x} + b = 0$.
3. Identify the support vectors in the dataset.
4. Calculate the margin, which is the perpendicular distance from the hyperplane to the nearest support vector.
5. Classify a new data point, $(3, 2)$, based on the learned SVM model.

3 Logistic Regression

You are working on a binary classification problem where you are using logistic regression to predict whether a student will pass (1) or fail (0) an exam based on the number of hours they have studied. You have collected the data in the provided table.

Hours Studied (X)	Pass (Y)
3	0
4	0
5	1
6	0
7	1
8	1
9	1

You want to fit a logistic regression model to this data and find the best-fitting sigmoid function, which is represented as:

$$P(Y = 1|X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X)}}$$

Where $P(Y = 1|X)$ is the probability of passing the exam given the hours studied.

1. Calculate the probabilities $P(Y = 1|X)$ for each data point using the logistic regression model.
2. Calculate the log-odds for each data point, which is $\log\left(\frac{P(Y=1|X)}{1-P(Y=1|X)}\right)$.
3. Calculate the cost function (log loss) for the given data and the predicted probabilities. The log loss for a single data point is given as:

$$\text{Log Loss} = -\left(Y \cdot \log(P(Y = 1|X)) + (1 - Y) \cdot \log(1 - P(Y = 1|X))\right)$$

4. Calculate the total cost (log loss) for all data points.
5. Use gradient descent or any other optimization method to find the values of β_0 and β_1 that minimize the total cost.