`http://ml3.leuphana.de/lectures/summer24/DL`
Machine Learning Group, Leuphana University of Lüneburg

Soham Majumder (soham.majumder@leuphana.de)

Classwork

Question 1

Consider the univariate function

$$f(x) = x^3 + 6x^2 - 3x - 5$$

Find its stationary points and indicate whether they are maximum, minimum, or saddle points.

Solution

Given the function $f(x)$, we obtain the following gradient and Hessian,

$$\frac{df}{dx} = 3x^2 + 12x - 3$$
$$\frac{d^2 f}{dx^2} = 6x + 12.$$

We find stationary points by setting the gradient to zero, and solving for x. One option is to use the formula for quadratic functions, but below we show how to solve it using completing the square. Observe that

$$(x + 2)^2 = x^2 + 4x + 4$$

and therefore (after dividing all terms by 3),

$$\frac{x^2 + 4x}{3} - 1 = \big((x + 2)^2 - 4\big) - 1.$$

By setting this to zero, we obtain that

$$(x + 2)^2 = 5,$$

and hence

$$x = -2 \pm \sqrt{5}.$$

Substituting the solutions of $\frac{df}{dx} = 0$ into the Hessian gives

$$\frac{d^2 f}{dx^2}(-2-\sqrt{5}) \approx -13.4 \quad \text{and} \quad \frac{d^2 f}{dx^2}(-2+\sqrt{5}) \approx 13.4.$$

This means that the left stationary point is a maximum and the right one is a minimum. See Figure 1 for an illustration.

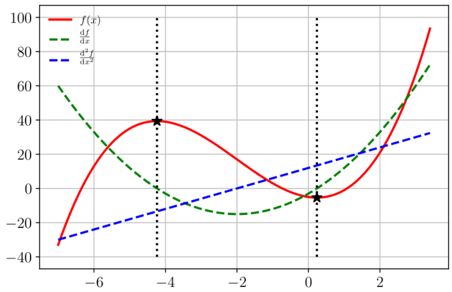

Figure 1: A plot of the function $\mathbf{f}(\mathbf{x})$ with its gradient and Hessian.

Question 2

Consider the following convex optimization problem

$$\begin{aligned} \min_{\mathbf{w}\in\mathbb{R}^D} \quad & \frac{1}{2}\mathbf{w}^T\mathbf{w} \\ \text{subject to} \quad & \mathbf{w}^T\mathbf{x} \geq 1. \end{aligned}$$

Derive the Lagrangian dual by introducing the Lagrange multiplier λ.

Solution

First, we express the convex optimization problem in standard form,

$$\begin{aligned} \min_{\mathbf{w}\in\mathbb{R}^D} \quad & \frac{1}{2}\mathbf{w}^T\mathbf{w} \\ \text{subject to} \quad & 1-\mathbf{w}^T\mathbf{x} \leq 0. \end{aligned}$$

By introducing a Lagrange multiplier $\lambda \geq 0$, we obtain the following Lagrangian

$$\mathcal{L}(\mathbf{w}) = \frac{1}{2}\mathbf{w}^T\mathbf{w} + \lambda(1-\mathbf{w}^T\mathbf{x})$$

Taking the gradient of the Lagrangian with respect to $\mathbf{w}$ gives

$$\frac{d\mathcal{L}(\mathbf{w})}{d\mathbf{w}} = \mathbf{w} - \lambda\mathbf{x}^T.$$

Setting the gradient to zero and solving for $\mathbf{w}$ gives

$$\mathbf{w} = \lambda \mathbf{x}.$$

Substituting back into $\mathcal{L}(\mathbf{w})$ gives the dual Lagrangian

$$\mathcal{D}(\lambda) = \frac{\lambda^2}{2}\mathbf{x}^T\mathbf{x} + \lambda - \frac{\lambda^2}{2}\mathbf{x}^T\mathbf{x} = -\frac{\lambda^2}{2}\mathbf{x}^T\mathbf{x} + \lambda.$$

Therefore, the dual optimization problem is given by

$$\begin{aligned} \max_{\lambda \in \mathbb{R}} \quad & -\frac{\lambda^2}{2}\mathbf{x}^T\mathbf{x} + \lambda \\ \text{subject to} \quad & \lambda \geq 0. \end{aligned}$$

Question 3

Show that changing the condition $y_n(\mathbf{w}^\top \mathbf{x}_n + b) \geq 1$ in SVM to a different condition $y_n(\mathbf{w}^\top \mathbf{x}_n + b) \geq m$ does not change the effective separating hyperplane that the SVM learns. Assume the hard-margin SVM for simplicity.

Solution

The original hard-margin SVM optimization problem can be stated as

$$\begin{aligned} \arg\min_{\mathbf{w}, b} \quad & \frac{1}{2}\|\mathbf{w}\|^2 \\ \text{subject to} \quad & y_n(\mathbf{w}^\top \mathbf{x}_n + b) \geq 1, \quad \forall n = 1, 2, ..., N \end{aligned}$$

The modified version of SVM involves changing the inequalities as $y_n(\mathbf{w}^\top \mathbf{x}_n + b) \geq m$, $\forall n = 1, 2, ..., N$. Let $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, ..., \alpha_N]$ be the Lagrange variables. The Lagrangian can be stated as

$$L(\mathbf{w}, b, \boldsymbol{\alpha}) = \frac{1}{2}\|\mathbf{w}\|^2 + \sum_{n=1}^{N} \alpha_n (m - y_n(\mathbf{w}^\top \mathbf{x}_n + b))$$

Using the dual formulation to solve the constrained optimization problem, we have

$$\frac{\partial L(\mathbf{w}, \boldsymbol{\alpha})}{\partial \mathbf{w}} = \mathbf{0} \implies \mathbf{w} = \sum_{n=1}^{N} \alpha_n y_n \mathbf{x}_n, \quad \frac{\partial L(\mathbf{w}, \boldsymbol{\alpha})}{\partial b} = 0 \implies \sum_{n=1}^{N} \alpha_n y_n = 0$$

Now, substituting $\mathbf{w} = \sum_{n=1}^{N} \alpha_n y_n \mathbf{x}_n$ in Lagrangian, we get the dual problem as

$$L_D(\boldsymbol{\alpha}) = \mathbf{w}^\top \mathbf{w} + \sum_{n=1}^{N} (\alpha_n m - y_n b) - \sum_{n=1}^{N} y_n \mathbf{w}^\top \mathbf{x}_n - \frac{1}{2} \sum_{n,l=1}^{N} \alpha_n \alpha_l y_n y_l (\mathbf{x}_n^\top \mathbf{x}_l)$$

Now, the objective can be stated in a compact form as

$$\max_{\boldsymbol{\alpha} \geq 0} m \cdot (\mathbf{1}^\top \boldsymbol{\alpha}) - \frac{1}{2} \boldsymbol{\alpha}^\top G \boldsymbol{\alpha}$$

where G is an $N \times N$ matrix with $G_{nl} = y_n y_l \mathbf{x}_n^\top \mathbf{x}_l$, and $\mathbf{1}$ is a vector of ones. Note that m simply turns out to be a multiplicative constant and thus does not affect the solution of the optimization problem. Thus, the solution to the modified SVM effectively remains the same as that of the original one.