CS282 Machine Learning: Quiz 1

CS282 TA Group
SIST
ShanghaiTech University

Problem 1

Given functions $g(\boldsymbol{x}) = \|\mathbf{Q}\boldsymbol{x} - \boldsymbol{b}\|_2^2$ and $f(\boldsymbol{x}) = \boldsymbol{x}^T\mathbf{Q}\boldsymbol{x} + \boldsymbol{b}^T\boldsymbol{x}$, $\mathbf{Q} \in \mathbb{R}^{n\times n}$ (is symmetric). Write down the gradients and Hessian.

Solution. Note that $\nabla_{\boldsymbol{x}}\left(\boldsymbol{x}^T\mathbf{A}\boldsymbol{x}\right) = 2\mathbf{A}\boldsymbol{x}$, $\nabla_{\boldsymbol{x}}\left(\boldsymbol{b}^T\boldsymbol{x}\right) = \boldsymbol{b}$, and $\nabla_{\boldsymbol{x}}^2\left(\boldsymbol{x}^T\mathbf{A}\boldsymbol{x}\right) = 2\mathbf{A}$, we have:

$$
\begin{aligned}
\nabla_{\boldsymbol{x}} g(\boldsymbol{x}) &= 2\mathbf{Q}^T(\mathbf{Q}\boldsymbol{x} - \boldsymbol{b}), & \nabla_{\boldsymbol{x}}^2 g(\boldsymbol{x}) &= 2\mathbf{Q}^T\mathbf{Q} \\
\nabla_{\boldsymbol{x}} f(\boldsymbol{x}) &= 2\mathbf{Q}\boldsymbol{x} + \boldsymbol{b} & \nabla_{\boldsymbol{x}}^2 f(\boldsymbol{x}) &= 2\mathbf{Q}
\end{aligned}
$$

Remark Note that for general $\mathbf{A} \in \mathbb{R}^{n\times n}$, we have $\nabla_{\boldsymbol{x}}\left(\boldsymbol{x}^T\mathbf{A}\boldsymbol{x}\right) = (\mathbf{A}^T + \mathbf{A})\boldsymbol{x}$, but usually we assume that the matrix $\mathbf{A}$ in the bilinear-form $\boldsymbol{x}^T\mathbf{A}\boldsymbol{x}$ is symmetric.

Problem 2

For symmetric $\mathbf{Q} \in \mathbb{R}^{n\times n}$, give the definition of positive definiteness.

What is the value of λ guaranteeing the positive definiteness of $\lambda\mathbf{I} + \mathbf{Q}^T\mathbf{Q}$.

Solution. Positive definiteness A symmetric matrix $\mathbf{Q} \in \mathbb{R}^{n\times n}$ is positive definite if

$$\forall \boldsymbol{x} \in \mathbb{R}^n, \boldsymbol{x} \neq 0, \boldsymbol{x}^T\mathbf{Q}\boldsymbol{x} > 0. \tag{1}$$

For a given symmetric matrix $\mathbf{Q} \in \mathbb{R}^{n\times n}$, by the definition of positive definiteness,

$$
\begin{aligned}
\lambda\mathbf{I} + \mathbf{Q}^T\mathbf{Q} \text{ is positive defnite} &\Leftrightarrow \boldsymbol{x}^T(\lambda\mathbf{I} + \mathbf{Q}^T\mathbf{Q})\boldsymbol{x} > 0, \forall \boldsymbol{x} \neq 0 \\
&\Leftrightarrow \lambda > -\frac{\boldsymbol{x}^T\mathbf{Q}^T\mathbf{Q}\boldsymbol{x}}{\boldsymbol{x}^T\boldsymbol{x}}, \forall \boldsymbol{x} \neq 0 \\
&\Leftrightarrow \lambda > \max_{\boldsymbol{x}\neq 0} -\frac{\boldsymbol{x}^T\mathbf{Q}^T\mathbf{Q}\boldsymbol{x}}{\boldsymbol{x}^T\boldsymbol{x}} \\
&\Leftrightarrow \lambda > -\min_{\boldsymbol{x}\neq 0} \frac{\boldsymbol{x}^T\mathbf{Q}^T\mathbf{Q}\boldsymbol{x}}{\boldsymbol{x}^T\boldsymbol{x}} \\
&\Leftrightarrow \lambda > -\lambda_{\min}(\mathbf{Q}^T\mathbf{Q})
\end{aligned}
$$

where $\lambda_{\min}(\mathbf{Q}^T\mathbf{Q})$ is the minimum eigenvalue of $\mathbf{Q}^T\mathbf{Q}$. The last inequality follows from Rayleigh quotient.

If we consider all symmetric matrix $\mathbf{Q} \in \mathbb{R}^{n\times n}$, we need to take an upper bound of $-\lambda_{\min}(\mathbf{Q}^T\mathbf{Q})$, which is 0. Thus we require that $\lambda > 0$.

Problem 3

For closed set $C \subset \mathbb{R}^n$, give the definition of convex set.

For function $f : \mathbb{R}^n \to \mathbb{R}$, give the definition of convex functions.

Solution. **convex set**: C is convex if $\forall \boldsymbol{x}, \boldsymbol{y} \in C, \forall \theta \in [0, 1]$, we have

$$\theta \boldsymbol{x} + (1-\theta)\boldsymbol{y} \in C \tag{2}$$

convex function: f is convex if:

1. $\mathrm{dom}(f)$ is convex.
2. $\forall \boldsymbol{x}, \boldsymbol{y} \in \mathrm{dom}(f), \forall \theta \in [0, 1]$, we have

$$f\left(\theta \boldsymbol{x} + (1-\theta)\boldsymbol{y}\right) \leq \theta f(\boldsymbol{x}) + (1-\theta) f(\boldsymbol{y}) \tag{3}$$

Remark It is not enough to say that f is convex if $\forall \boldsymbol{x}, \boldsymbol{y} \in \mathrm{dom}(f)$, we have

$$f\left(\frac{\boldsymbol{x}+\boldsymbol{y}}{2}\right) \leq \frac{1}{2}\left(f(\boldsymbol{x}) + f(\boldsymbol{y})\right) \tag{4}$$

since this condition requires that f is continuous. See Midpoint-convex doesn't imply convex.

Problem 4

Write the first 4 terms of the Taylor expansion of $f(x) = e^x + 3x + (x-1)^2$ at $\hat{x} = 0$.

Write the first two terms of Taylor expansion of $f : \mathbb{R}^n \to \mathbb{R}$ at $\hat{\boldsymbol{x}}$.

Solution. Expanding $f(x)$ yields:

$$f(x) = \left(1 + x + \frac{x^2}{2} + \frac{x^3}{6} + o(x^3)\right) + 3x + (x-1)^2 = 2 + 2x + \frac{3x^2}{2} + \frac{x^3}{6} + o(x^3) \tag{5}$$

Note that f is a multi-variable function, of which Taylor approximation is given by

$$f(\boldsymbol{x}) \approx f(\hat{\boldsymbol{x}}) + \nabla f(\hat{\boldsymbol{x}})^T (\boldsymbol{x} - \hat{\boldsymbol{x}}) \tag{6}$$

Problem 5

Consider two **independent** random variables X and Y. X is continuous with probability density function $f(x)$ for $x \in \mathbb{R}$. Y is a Bernoulli distribution with probability mass function $g(y)$ for $y \in [0, 1]$. Write the formulation of the expectation of random variable $\phi(X, Y)$

Solution. Suppose the pdf of joint probability distribution of X, y is $h(X, Y)$, since X and Y are independent variables, we have

$$h(X = x, Y = i) = g(y = i) f(x)$$

Thus we have

$$\begin{aligned}
\mathbb{E}\left[\phi(X, Y)\right] &= \sum_{i=0,1} \int_{\mathbb{R}} \phi(x, y = i) h(x, y = i) dx \\
&= \sum_{i=0,1} \int_{\mathbb{R}} \phi(x, y = i) g(y = i) f(x) dx \\
&= g(0) \int_{\mathbb{R}} \phi(x, 0) f(x) dx + g(0) \int_{\mathbb{R}} \phi(x, 1) f(x) dx
\end{aligned}$$

References

[1] Petersen, K. B. and Pedersen, M. S. The Matrix Cookbook. , Technical University of Denmark (2008). , Version 20081110 .