Homework 2 ML

Shreyas Shukla

Question 4

1. Linear: $f(x) = y^T x$

$$\frac{d}{dx}(y^T x) = \frac{df}{dx} = \begin{bmatrix} \frac{df}{dx_1}(x) \\ \frac{df}{dx_2}(x) \\ \vdots \\ \frac{df}{dx_n}(x) \end{bmatrix} = y$$

2. Quadratic: $f(x) = x^T A x$

$$\frac{df}{dx} = 2Ax$$

3. Trace of Quadratic: $f(x) = \mathbf{Tr}(x^T A x)$

$$df = d(\mathrm{Tr}(x^T A x))$$

$$= \mathrm{Tr}(d(x^T)Ax) + \mathrm{Tr}(x^T A d(x))$$

$$= \mathrm{Tr}(dx^T A x) + \mathrm{Tr}(x^T A dx)$$

$$= x^T A dx + dx^T A x$$

$$= 2Ax$$

4. Activation Function: $f(w) = \mathbf{ReLU}(w^T x)$

$$\frac{df}{dw} = \begin{cases} x & \text{if } w^T x > 0 \\ 0 & \text{otherwise} \end{cases}$$

5. Linear Regression: $f(w) = \frac{1}{2}\sum_{i=1}^{n}(w^T x_i - y_i)^2$

$$\frac{df}{dw} = \sum_{i=1}^{n}(w^T x_i - y_i)x_i$$

6. Multivariate Gaussian Tr: $f(x) = e^{-\mathbf{Tr}(x^T Ax)}$

$$df = d\left(e^{-\mathrm{Tr}(x^T Ax)}\right)$$

$$df = e^{-\mathrm{Tr}(x^T Ax)} \cdot d\left(-\mathrm{Tr}(x^T Ax)\right)$$

$$d\left(\mathrm{Tr}(x^T Ax)\right) = 2Ax$$

$$df = -2Axe^{-\mathrm{Tr}(x^T Ax)}$$

7. Multivariate Gaussian: $f(x) = e^{-x^T Ax}$

$$df = d\left(e^{-x^T Ax}\right)$$

$$df = e^{-x^T Ax} \cdot d\left(-x^T Ax\right)$$

$$d\left(x^T Ax\right) = 2Ax$$

$$df = -2Axe^{-x^T Ax}$$

8. Sigmoid Function: $f(z) = \frac{1}{1+e^{-z}}$

$$f(z) = \frac{1}{1+e^{-z}}$$

$$f(z) = (1+e^{-z})^{-1}$$

$$\frac{df}{dz} = \frac{d}{dz}\left((1+e^{-z})^{-1}\right)$$

$$\frac{df}{dz} = -(1+e^{-z})^{-2} \cdot \frac{d}{dz}(1+e^{-z})$$

$$\frac{d}{dz}(1+e^{-z}) = -e^{-z}$$

$$\frac{df}{dz} = \frac{e^{-z}}{(1+e^{-z})^2}$$

$$\frac{df}{dz} = f(z)(1-f(z))$$

9. Logistic Regression Objective: $f(w) = \frac{1}{1+e^{-w^T x}}$

$$f(w) = \frac{1}{1+e^{-w^T x}}$$

$$f(w) = (1+e^{-w^T x})^{-1}$$

$$\frac{df}{dw} = \frac{d}{dw}\left((1+e^{-w^T x})^{-1}\right)$$

$$\frac{df}{dw} = -(1+e^{-w^T x})^{-2} \cdot \frac{d}{dw}(1+e^{-w^T x})$$

$$\frac{d}{dw}(1+e^{-w^T x}) = -e^{-w^T x}x$$

$$\frac{df}{dw} = \frac{e^{-w^T x}x}{(1+e^{-w^T x})^2}$$

$$\frac{df}{dw} = f(w)(1-f(w))x$$

10. L1 Norm: $f(x) = \|x\|_1$

$$\frac{df}{dx} = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x = 0 \\ -1 & \text{if } x < 0 \end{cases}$$

11. L2 Norm: $f(x) = \|x\|_2$

$$f(x) = \|x\|_2 = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2} = \sqrt{x^T x}$$

$$f(x) = (x^T x)^{1/2}$$

$$\frac{df}{dx} = \frac{1}{2}(x^T x)^{-1/2} \cdot \frac{d}{dx}(x^T x)$$

$$\frac{d}{dx}(x^T x) = 2x$$

$$\frac{df}{dx} = \frac{1}{2}(x^T x)^{-1/2} \cdot 2x$$

$$\frac{df}{dx} = \frac{x}{\sqrt{x^T x}} = \frac{x}{\|x\|_2}$$

12. SVM Objective: $f(w) = \sum_{i=1}^{n} \textbf{ReLU}(1 - y_i(x_i, w))$

$$\frac{df}{dw} = -\sum_{i=1}^{n} y_i x_i \quad \text{if } 1 - y_i(x_i, w) > 0$$

13. PCA Objective: $f(x) = x^T A x - \lambda(x^T x - 1)$

$$\frac{df}{dx} = 2Ax - 2\lambda x$$

14. Gaussian Maximum Likelihood:

$$f(\mu) = \log\left(\prod_i \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i - \mu)^2}{2\sigma^2}}\right)$$

$$\frac{df}{d\mu} = \sum_i \frac{x_i - \mu}{\sigma^2}$$

15. Exponential Maximum Likelihood:

$$f(\lambda) = \log\left(\prod_i \lambda e^{-\lambda x_i}\right)$$

$$\frac{df}{d\lambda} = \frac{n}{\lambda} - \sum_i x_i$$

16. Bernoulli Maximum Likelihood:

$$f(p) = \log\left(\prod_i p^{\alpha_i} (1 - p)^{1 - \alpha_i}\right)$$

$$\frac{df}{dp} = \sum_i \frac{\alpha_i}{p} - \frac{1 - \alpha_i}{1 - p}$$

17. Uniform Maximum Likelihood:

$$f(a,b) = \log\left(\prod_i \frac{1}{b-a}\right)$$

$$\frac{df}{da} = -\frac{n}{b-a}$$

$$\frac{df}{db} = \frac{n}{b-a}$$