Solutions Sheet

Nina Fischer and Yannick Zelle

November 16, 2021

Exercise 1

(a)

Let $x, b \in \mathbb{R}^m, \alpha \in \mathbb{R}$ and $A \in R^{mxm}$. We want to find the gradient of :

$$\psi = x^T Ax + b^T x + \alpha$$

We have:

$$\begin{aligned}
d\psi &= d(x^T Ax + b^T x + \alpha) \\
&= dx^T Ax + db^T x + d\alpha \\
&= dx^T Ax + db^T x \\
&= x^T (A + A^T)dx + db^T x \\
&= x^T (A + A^T)dx + d(b)^T x + b^T dx \\
&= (x^T (A + A^T) + b^T)dx
\end{aligned}$$

So:

$$D\psi = x^T (A + A^T) + b^T$$

(b)

Let $x_i \in \mathbb{R}^n$ and $y_i \in \{0, 1\}$ for $1 \leq i \leq n$ We are looking for the dervative of

$$\tau = \sum_{i=1}^{n} (y_i - \sigma(x_i^T \omega + b))^2$$

,where:

$$\sigma(\alpha) = \frac{1}{1 + exp(\alpha)}$$

We have:

$$\begin{aligned}
d\tau &= d(\sum_{i=1}^{n}(y_i - \sigma(x_i^T\omega + b))^2) \\
&= \sum_{i=1}^{n} d((y_i - \sigma(x_i^T\omega + b))^2) \\
&= \sum_{i=1}^{n} 2(y_i - \sigma(x_i^T\omega + b))d(y_i - \sigma(x_i^T\omega + b)) \\
&= \sum_{i=1}^{n} 2(y_i - \sigma(x_i^T\omega + b))(dy_i - d\sigma(x_i^T\omega + b)) \\
&= -2\sum_{i=1}^{n}(y_i - \sigma(x_i^T\omega + b))(\sigma(x_i^T\omega + b) \cdot (1 - \sigma(x_i^T\omega + b))d(x_i^T\omega + b) \\
&= -2\sum_{i=1}^{n}(y_i - \sigma(x_i^T\omega + b))(\sigma(x_i^T\omega + b) \cdot (1 - \sigma(x_i^T\omega + b))db
\end{aligned}$$

So the dervative with respect to b is:

$$D_b\tau = -2\sum_{i=1}^{n}(y_i - \sigma(x_i^T\omega + b))(\sigma(x_i^T\omega + b) \cdot (1 - \sigma(x_i^T\omega + b))$$

And for ω we have :

$$\begin{aligned}
d\tau &= -2\sum_{i=1}^{n}(y_i - \sigma(x_i^T\omega + b))(\sigma(x_i^T\omega + b) \cdot (1 - \sigma(x_i^T\omega + b))d(x_i^T\omega + b) \\
&= -2\sum_{i=1}^{n}(y_i - \sigma(x_i^T\omega + b))(\sigma(x_i^T\omega + b) \cdot (1 - \sigma(x_i^T\omega + b))x_i^T d\omega
\end{aligned}$$

So we have :

$$D_b\omega = -2\sum_{i=1}^{n}(y_i - \sigma(x_i^T\omega + b))(\sigma(x_i^T\omega + b) \cdot (1 - \sigma(x_i^T\omega + b))x_i^T$$

Let $x, y \in \mathbb{R}^m, A, B \in \mathbb{R}^{mxm}$ and σ the sigmoid function we are lookig for the gradient of :

$$\psi(x) = ||y - A\sigma(Bx)||_2^2$$

(c)

let $x, y \in \mathbb{R}^m, A, B \in \mathbb{R}^{mxm}$ and σ the sigmoid function. We want to find the gradinet of

$$\psi = ||y - A\sigma(Bx)||_2^2$$

By using the definition of the euklidean Norm we have:

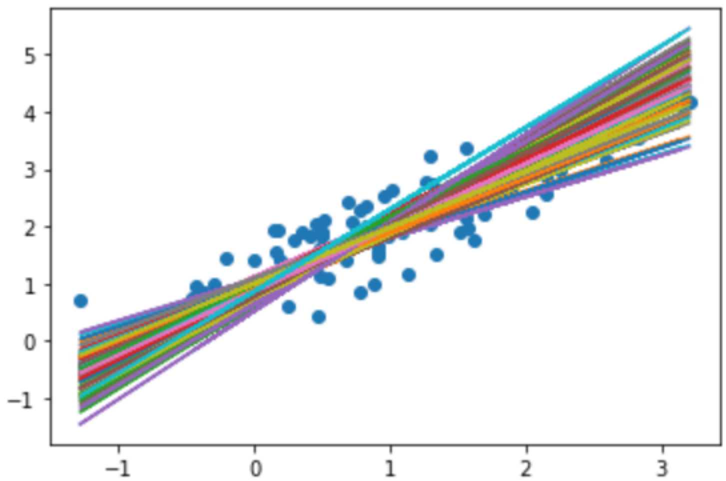
$$\begin{aligned}
d\psi &= d||y - A\sigma(Bx)||_2^2 \\
&= d((\sum_{i=1}^{m}(y_i - \sigma(Bx)_i)^2)^{\frac{1}{2}})^2 \\
&= d\sum_{i=1}^{m}(y_i - \sigma(Bx)_i)^2 \\
&= \sum_{i=1}^{m} d(y_i - \sigma(Bx)_i)^2 \\
&= \sum_{i=1} 2(y_i - \sigma(Bx)_i)d(y_i - \sigma(Bx)_i) \\
&= \sum_{i=1} 2(y_i - \sigma(Bx)_i)(dy_i - d\sigma(Bx)_i) \\
&= \sum_{i=1} 2(y_i - \sigma(Bx)_i)(-d\sigma(Bx)_i) \\
&= \sum_{i=1} 2(y_i - \sigma(Bx)_i)(-\sigma(Bx)_i \odot (1 - -\sigma(Bx)_i))dBx_i \\
&= \sum_{i=1} 2(y_i - \sigma(Bx)_i)(-\sigma(Bx)_i \odot (1 - -\sigma(Bx)_i))Bdx_i
\end{aligned}$$

Then we have :

$$D\psi = \sum_{i=1} 2(y_i - \sigma(Bx)_i)(-\sigma(Bx)_i \odot (1 - -\sigma(Bx)_i))Bdx_i$$

Exercise 2

See Jupyter Notebook for details...



Exercise 3

Exercise 4