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 :

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

(c)

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

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

By using the definition of the Euclidean 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 :

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

(d)

We want to proof that $d(\phi^\alpha) = \alpha\phi^{\alpha-1}d\phi$ by induction.

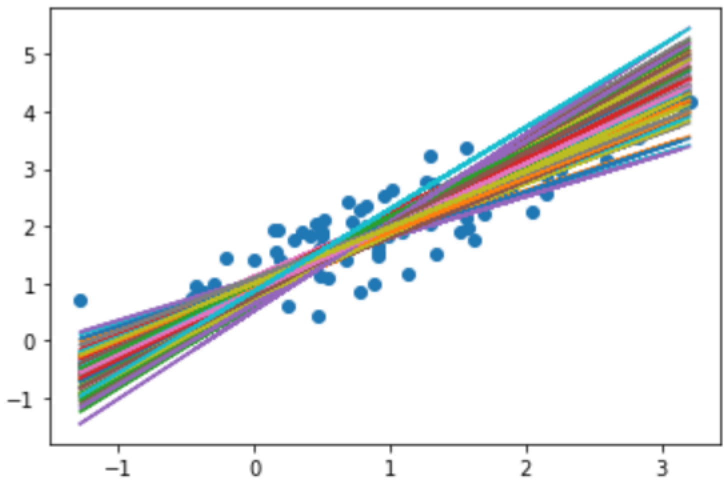
For $\alpha = 1$ we have $d(\phi^1 = d\phi = 1\phi^0 d\phi = d\phi$.

For $\alpha \to \alpha + 1$ we want to prove that $d(\phi^{\alpha+1}) = (\alpha + 1)\phi^{(\alpha+1)-1}d\phi$.

$$
\begin{aligned}
d(\phi^{\alpha+1}) &= d(\phi^\alpha\phi) \\
&= d(\phi^\alpha)\phi + \phi^\alpha d(\phi) \\
&= \alpha\phi^{\alpha-1}\phi d\phi + \phi^\alpha d\phi \\
&= (\alpha\phi^{\alpha-1}\phi + \phi^\alpha)d\phi \\
&= (\alpha + 1)\phi^\alpha d\phi
\end{aligned}
$$

Exercise 2

See Jupyter Notebook for details...



Exercise 3

We want to calculate the derivative of $p(y|X,\omega) = \mathcal{N}(y|X\omega, \sigma^2 I)$.

We take $a = \frac{1}{(2\pi)\det(\sigma^2 I)}$ and $b = exp(-\frac{1}{2}(y - X\omega)^T \sigma^2 I^{-1}(y - X\omega))$.

$$
\begin{aligned}
d(p(y|X,\omega) &= d\mathcal{N}(y|X\omega, \sigma^2 I) \\
&= d\frac{1}{(2\pi)\det(\sigma^2 I)} exp(-\frac{1}{2}(y - X\omega)^T \sigma^2 I^{-1}(y - X\omega) \\
&= adexp(-\frac{1}{2}(y - X\omega)^T \sigma^2 I^{-1}(y - X\omega))exp(-\frac{1}{2}(y - X\omega)^T \sigma^2 I^{-1}(y - X\omega)) \\
&= a(d - \frac{1}{2}(y - X\omega)^T \sigma^2 I^{-1}(y - X\omega) - \frac{1}{2}(y - X\omega)^T d\sigma^2 I^{-1}(y - X\omega))b \\
&= a(-\frac{1}{2} - dX\omega)^T \sigma^2 I^{-1}(y - X\omega) - \frac{1}{2}(y - X\omega)^T \sigma^2 I^{-1}(-dX\omega))b \\
&= a(-\frac{1}{2}(-1)((dX)\omega + Xd\omega)\sigma^2(y - X\omega) - \frac{1}{2}(y - X\omega)^T \sigma^2 I^{-1}(-(dX)\omega + Xd\omega))b \\
&= a(\frac{1}{2}Xd\omega\sigma^2(y - X\omega) - \frac{1}{2}(y - X\omega)^T \sigma^2 I^{-1} Xd\omega)b \\
&= a(\frac{1}{2}X\sigma^2(y - X\omega) - \frac{1}{2}(y - X\omega)^T \sigma^2 I^{-1} X)bd\omega
\end{aligned}
$$