ADVANCED COURSE IN MACHINE LEARNING: 1

(1) frobenius norm for a IXI matrix:
$$\|X\|^2 = \frac{1}{2} \sum_{i=1}^{2} X_{ij}^2 = \text{trace}(X^T X)$$

a) and the derivative is $\frac{d}{dX} \|X\|^2 = 2X$. Prove it.

Now we fix (iii) =
$$\frac{d+(x)}{dxij} = 2xij$$
 = $\frac{d+(x)}{dx} = 2x$

because $f(X) = X_{11}^2 + X_{12}^2 + ... + X_{13}^2 + ... + X_{13}^2$ so it we this (i,j) the derivative will be 2Xij.

- b) let $f(x) = (x+a)^T A^{-1}(Bx)$ where x is a column vector $x = x^{1\times D}$. Compute the gradient $\nabla_x f(x)$.
 - · X is a vector column DXI and a is the same.
 - · (x+a) is a row vector 1.x D
 - · A.B are square matrices DxD (IR DXD)
 - · Also (x+a) TA-1 is a now vector 1 x D (IR2x0)
 - · Bx is a column vector Dx1
 - · Finally +(x) & IR

PROPERTIES:
$$\frac{d(x^TBx)}{dx} = x^T(B+B^T)$$
 and $\frac{d(a^Tx)}{dx} = a^T$

so applying them: $\nabla_x + (x) = X^T (A^{-1}B + (A^{-1}B)^T) + \alpha^T A^{-1}B$ Notice that $\nabla_x + (x) \in \mathbb{R}^{1 \times 10}$ so it is a now vector.

- (2) N data points $Xn \in IR^{D\times 1}$ with autputs yn. Predict $\overline{y}_n = e^{XnO} + b$ where $O \in IR^{D\times 1}$ and $b \in IR$.

 We want to minimize the squared errors: $L = \sum_{n=1}^{N} w_n (y_n \overline{y}_n)^2$ for $w_n > 0$.
 - a) Write the loss in matrix notation.

$$X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} x_{11} \dots x_{1D} \\ \vdots \\ x_{n1} \dots x_{nD} \end{pmatrix} \qquad Y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \qquad X \in \mathbb{R}^{n \times D}$$

$$Y = \begin{pmatrix} \overline{y_1} \\ \vdots \\ \overline{y_n} \end{pmatrix} = \begin{pmatrix} e^{x_1^T \phi + b} \\ \vdots \\ e^{x_n^T \phi + b} \end{pmatrix} = e^{x_1^T \phi + b}$$

$$\overline{Y} = \begin{pmatrix} \overline{y_1} \\ \vdots \\ \overline{y_n} \end{pmatrix} = \begin{pmatrix} e^{x_1^T \phi + b} \\ \vdots \\ e^{x_n^T \phi + b} \end{pmatrix} = e^{x_n^T \phi + b}$$

Then the loss: L = 11 W2(Y-Y) 112 = 11 W2(Y-exo+b) 112

b) Derivative wit @ and b.

$$\frac{dL}{d\theta} = x \left(-e^{x\theta+b}\right) W 2 \left(W \left(y - e^{x\theta+b}\right)\right)$$

$$\frac{dL}{dh} = \left(-e^{x\theta+b}\right) W 2 \left(W \left(y - e^{x\theta+b}\right)\right)$$

c) The gradient

our founds, a trac ded times of

(ara y and sadroon should su a his

(Brain) and were not a statement with

CACE TOLDER OFFILE BUILD

A B A C TOLL A STALL A

Action of the state of the stat

у С. — у голы селофи зан озатилир на техно м Сту

nongalan seriah se senerah sebias se

and the track of the same and the same