Advanced Topics in Statistical Machine Learning: Assignment 1

Mateus Zitelli Dantas

28/06/2016

1 Question 1

Proof by induction of $det(A) = det(A^T)$ where $A \in \mathbb{R}^{n \times n}$.

1.1 Base case with $n = 1$:

$$A = A^T \implies det(A) = det(A^T)$$

1.2 Induction hypothesis:

$$det(M) = det(M^T) \; \forall M \in \mathbb{R}^{n \times n}$$

1.3 Induction step:

Suppose $B = A^T | A, B \in \mathbb{R}^{n+1 \times n+1}$. Therefore $a_{i,j} = b_{j,i}$ and $B_{j,i} = A^T_{i,j}$ where $M_{i,j}$ is a minor matrix. Using Laplace expansion the determinant of A can be expressed as:

$$det(A) = \sum_{j=1}^{n+1} (-1)^{i+j} a_{i,j} det(A_{i,j})$$

Therefore, considering the induction hypothesis we can write $det(A)$ as:

$$det(A) = \sum_{j=1}^{n+1} (-1)^{i+j} a_{i,j} det(A^T_{i,j})$$

Replacing with the elements of matrix B:

$$det(A) = \sum_{j=1}^{n+1} (-1)^{i+j} b_{j,i} det(B_{j,i}) = det(B) = det(A^T)$$

Q.E.D.

2 Question 2

Proof by induction of $det(I) = 1$ where I is the identity matrix.

2.1 Base case with for I^1:

Using Laplace expansion:

$$det(I) = 1$$

2.2 Induction hypothesis:

$$det(I^{n \times n}) = 1$$

2.3 Induction step:

Using Laplace expansion:

$$det(I^{n+1 \times n+1}) = \sum_{j=1}^{n+1} (-1)^{i+j} i_{i,j} det(I_{i,j})$$

Where $i_{i,j}$ is the element in the position i, j and $I_{i,j}$ is a minor matrix. Considering that we can write:

$$det(I^{n+1 \times n+1}) = (-1)^{i+i} i_{i,i} det(I_{i,i}) = det(I_{i,i})$$

And considering $I_{i,i} = I^{n \times n}$ and the induction hypothesis:

$$det(I^{n+1 \times n+1}) = 1$$

Q.E.D.

3 Question 3

First, the determinant of an triangular matrix can be expressed as:

$$det(D) = \prod_{n=1}^{n} d_{n,n}$$

And if D is a triangular matrix $\lambda I - D$ is also triangular. So, considering the characteristic equation $det(\lambda I - D) = 0$:

$$det(\lambda I - D) = \prod_{n=1}^{n} \lambda_i - d_{n,n} = 0 \implies \lambda_i = d_{i,i} \; \square$$

4 Question 4

Proof by counterexample:

$$A = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

5 Question 5: Gradients

To calculate matrix functions gradients we need to find the gradient to an arbitrary direction ξ using the derivative definition:

$$\mathcal{D}f(x)(\xi) = \lim_{h \to 0} \frac{f(x + h\xi) - f(x)}{h}$$

And then isolate such vector ξ in a inner product, resulting in $\langle \nabla_x f(x), \xi \rangle$.

5.1 a) $f(x) = a^T x + b$

$$\mathcal{D}f(x)(\xi) = \lim_{h \to 0} \frac{a^T(x + h\xi) + b - (a^T x + b)}{h} =$$
$$= a^T \xi = \langle a, \xi \rangle \implies \nabla_x f(x) = a$$

5.2 b) $f(x) = \frac{1}{2} x^T P x + q^T x + r$

$$\mathcal{D}f(x)(\xi) = \lim_{h \to 0} \frac{\frac{1}{2}(x + h\xi)^T P (x + h\xi) + q^T(x + h\xi) + r - (\frac{1}{2} x^T P x + q^T x + r)}{h} =$$
$$= q^T \xi + \frac{1}{2}(x^T P \xi + \xi^T P x)$$

Knowing that $\xi^T P x = (\xi^T P x)^T = x^T P^T \xi$ because it is an scalar:

$$\mathcal{D}f(x)(\xi) = q^T \xi + \frac{1}{2}(x^T P \xi + x^T P^T \xi) =$$

$$= (q^T + \frac{1}{2} x^T (P + P^T))\xi = \langle (q^T + \frac{1}{2} x^T (P + P^T))^T, \xi \rangle$$

Therefore:

$$\nabla_x f(x) = (q^T + \frac{1}{2} x^T (P + P^T))^T = q + \frac{1}{2}(P^T + P)x.$$

5.3 c) $f(x) = \frac{1}{2} x^T P x$

$$\mathcal{D}f(x)(\xi) = \frac{1}{2} \lim_{h \to 0} \frac{(x + h\xi)^T P (x + h\xi) - x^T P x}{h} = \frac{1}{2}(x^T P \xi + \xi^T P x)$$

Where $\xi^T P x = x^T P^T \xi = x^T P \xi$ because it is an scalar, therefore:

$$\mathcal{D}f(x)(\xi) = \frac{1}{2}(x^T P \xi + x^T P \xi) = x^T P \xi = \langle P^T x, \xi \rangle \implies \nabla_x f(x) = P^T x.$$

5.4 d) $f(x) = \frac{exp(a^Tx+b)}{1+exp(a^T+b)}$

$$\nabla_x f(x) = \frac{g'(x) + h(x) - g(x)h'(x)}{h(x)^2},$$

where

$$g(x) = exp(a^T x + b) \implies g'(x) = exp(a^T x + b)a$$
$$h(x) = 1 + exp(a^T x + b) \implies h'(x) = exp(a^T x + b)a$$

therefore:

$$\nabla_x f(x) = \frac{exp(a^T x + b)a}{(1 + exp(a^T x + b))^2}$$

.

6 Question 6

6.1 Symmetry

$$tr(A) = tr(A^T) \implies \langle X, Y\rangle = Tr(X^T Y) = Tr(Y^T X) = \langle Y, X\rangle.$$

6.2 Linearity

$$Tr(aB) = aTr(B) \implies \langle ax, y\rangle = Tr(aX^T Y) = a\ Tr(X^T Y) = a\ \langle X, Y\rangle.$$

$$Tr(A + B) = Tr(A) + Tr(B) \implies \langle X + Y, Z\rangle = Tr\{(X + Y)^T Z\} =$$

$$= Tr\{X^T Z + Y^T Z\} = Tr\{X^T Z\} + Tr\{Y^T Z\} = \langle X, Z\rangle + \langle Y, Z\rangle.$$

6.3 Positive-definiteness

$$Tr(X^T X) = \sum_{i,j} x_{i,j}^2 \geq 0$$

$$Tr(X^T X) = 0 \implies \sum_{i,j} x_{i,j}^2 = 0 \implies$$

$$\implies x_{i,j} = 0\ \forall (i, j) \in \{(i, j)|i \in \{1, ..., m\}, j \in \{1, ..., n\}\}.$$

7 Question 7

$x^T Ax$ is an scalar so:

$$x^T Ax = (x^T Ax)^T = x^T A^T x = -x^T Ax \implies x^T Ax = 0\ \square$$

8 Question 8

$$cov(X,Y)^2 \leq var(X)var(Y)$$

Considering $cov(X,Y]) = E[XY] - E[X]E[Y]$ and $var(X) = E[X^2] - E[X]^2$, therefore:

$$(E[XY] - E[X]E[Y])^2 \leq (E[X^2] - E[X]^2)(E[Y^2] - E[Y]^2) \implies$$

$$\implies E[XY]^2 \leq E[X^2]E[Y^2]$$

Expanding the expectations by the definition $E[f(x)] = \int f(x)P(x)dx$:

$$(\iint xyP(x,y)dxdy)^2 = \iint x^2y^2P(x,y)^2dxdy \leq \iint x^2y^2P(x)P(y)dxdy \implies$$

$$\implies \iint P(x,y)^2dxdy \leq \iint P(x)P(y)dxdy \implies$$

$$\implies P(x|y)^2P(y)^2 \leq P(x)P(y) \implies$$

$$\implies P(x|y)^2 \leq \frac{P(x)}{P(y)} \implies$$

$$\implies P(x|y)^2 \leq P(x|y)$$

Q.E.D.