COMP3670 Assignment 2

Ex 1. Orthogonal Projections.

Consider the Euclidean vector space $\mathbb{R}^3$ w/ the dot product. A subspace $U \subset \mathbb{R}^3$ and vector $x \in \mathbb{R}^3$ are given by:

$$U = \text{span}\left\{ \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ -1 \\ -2 \end{bmatrix} \right\}, \quad x = \begin{bmatrix} 8 \\ 4 \\ 16 \end{bmatrix}$$

1. Show that $x \notin U$.

Show x is not a linear combination of u_1 & u_2.

1. $-c_1 + 2c_2 = 8$
2. $c_1 - c_2 = 4$
3. $c_1 - 2c_2 = 16$

$$\begin{aligned} c_1 - c_2 &= 4 \\ - \quad c_1 - 2c_2 &= 16 \end{aligned} \quad = \quad \begin{aligned} c_2 &= -12 \\ c_1 &= -8. \end{aligned}$$

3. $-8 - 2 \cdot -12 = 16$
2. $-8 + 12 = 4$
1. $8 + (-24) \neq 8$

$\therefore x \notin U.$

2. Determine $\Pi_u(x)$.

$$\Pi_u(x) = \frac{\langle x, u_1 \rangle}{\|u_1\|^2} u_1 + \frac{\langle x, u_2 \rangle}{\|u^2\|^2} u_2$$

$$\langle x, u_1 \rangle = \begin{bmatrix} 8 \\ 4 \\ 16 \end{bmatrix} \cdot \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix} = -8+4+16 = 12$$

$$\langle x, u_2 \rangle = \begin{bmatrix} 8 \\ 4 \\ 16 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ -1 \\ -2 \end{bmatrix} = 16-4-32 = -20$$

$$\|u_1\|^2 = 1+1+1 = 3$$

$$\|u_2\|^2 = 4+1+4 = 9$$

$$\Pi_u(x) = \frac{12}{3} \cdot \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix} + \frac{-20}{9} \cdot \begin{bmatrix} 2 \\ -1 \\ -2 \end{bmatrix}$$

$$= \begin{bmatrix} -12/3 \\ 12/3 \\ 12/3 \end{bmatrix} + \begin{bmatrix} -40/9 \\ 20/9 \\ 40/9 \end{bmatrix}$$

$$\Pi_u(x) = \begin{bmatrix} -8.\overline{44} \\ 6.\overline{22} \\ 8.\overline{44} \end{bmatrix}$$

3. Determine the distance $d(x, U) := \min_{y \in U} \|x - y\|$

$$d(x, U) := \|x - \Pi_U(x)\| = \sqrt{\langle x - \Pi_U(x), x - \Pi_U(x) \rangle}$$

$$= \text{Sqrt}\left(\left(\begin{bmatrix} 8 \\ 4 \\ 16 \end{bmatrix} - \begin{bmatrix} -8.44 \\ 6.22 \\ 8.44 \end{bmatrix} \right) \cdot \left(\begin{bmatrix} 8 \\ 4 \\ 16 \end{bmatrix} - \begin{bmatrix} -8.44 \\ 6.22 \\ 8.44 \end{bmatrix} \right) \right)$$

$$= \text{Sqrt}\left(\begin{bmatrix} 16.44 \\ -2.22 \\ 7.56 \end{bmatrix} \cdot \begin{bmatrix} 16.44 \\ -2.22 \\ 7.56 \end{bmatrix} \right)$$

$$= \text{Sqrt}\left([270.2736 + 4.9284 + 57.1536] \right)$$

$$= \sqrt{332.3556} \approx 18.23.$$

4. Use Gram-Schmidt to transform $A = \begin{bmatrix} -1 & 2 \\ 1 & -1 \\ 1 & -2 \end{bmatrix}$ into B w/ orthonormal columns.

$$v_1 = a_1 = \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}$$

$$v_2 = a_2 - \frac{a_2 \cdot v_1}{v_1 \cdot v_1} \cdot v_1 = \begin{bmatrix} 2 \\ -1 \\ -2 \end{bmatrix} - \frac{-5}{3} \cdot \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1/3 \\ 2/3 \\ -1/3 \end{bmatrix}$$

Normalise the orthogonal basis to get orthonormal basis.

$$u_1 = \frac{v_1}{\|v_1\|}, \quad u_2 = \frac{v_2}{\|v_2\|}$$

$$u_1 = \begin{bmatrix} -0.577 \\ 0.577 \\ 0.577 \end{bmatrix}, \quad u_2 = \begin{bmatrix} 0.408 \\ 0.816 \\ -0.408 \end{bmatrix} \qquad B = \begin{bmatrix} -0.577 & 0.408 \\ 0.577 & 0.816 \\ 0.577 & -0.408 \end{bmatrix}$$

5. Find the vector θ that minimizes $\|x - Q\theta\|^2 + \lambda\|\theta\|^2$.

Differentiate the loss function w.r.t θ.

$F(\theta) = \|x - Q\theta\|^2 + \lambda\|\theta\|^2$

$F(\theta) = (x - Q\theta)^T(x - Q\theta) + \lambda\theta^T\theta$

expand:

$F(\theta) = (x^T - Q^T\theta^T)(x - Q\theta) + \lambda\theta^T\theta$

$F(\theta) = x^Tx - x^TQ\theta - xQ^T\theta^T + Q^T\theta^TQ\theta + \lambda\theta^T\theta$

$\frac{\partial F}{\partial \theta} = 0 - Q^Tx - Q^Tx + 2Q^TQ\theta + 2\lambda\theta$

$\frac{\partial F}{\partial \theta} = -2Q^Tx + 2Q^TQ\theta + 2\lambda\theta$

To minimise, set the gradient to zero

$0 = -2Q^Tx + 2Q^TQ\theta + 2\lambda\theta$

$0 = -Q^Tx + Q^TQ\theta + \lambda\theta$

$Q^Tx = Q^TQ\theta + \lambda\theta$

$Q^Tx = (Q^TQ + \lambda I)\theta$ since Q has orthonormal columns $Q^TQ = I$.

$Q^Tx = (I + \lambda I)\theta$

$Q^Tx = (1 + \lambda)\theta$

$\therefore \vec{\theta} = \frac{Q^Tx}{1 + \lambda}$.

6. Compute the vector θ for the matrix B & $\lambda = 10$.

$$B = \begin{bmatrix} -0.577 & 0.408 \\ 0.577 & 0.816 \\ 0.577 & -0.408 \end{bmatrix} \qquad x = \begin{bmatrix} 8 \\ 4 \\ 16 \end{bmatrix}$$

$$\theta = \frac{B^T x}{1+\lambda} = \frac{\begin{bmatrix} -0.577 & 0.577 & 0.577 \\ 0.408 & 0.816 & -0.408 \end{bmatrix} \cdot \begin{bmatrix} 8 \\ 4 \\ 16 \end{bmatrix}}{11}$$

$$\theta = \frac{\begin{bmatrix} -0.577 \times 8 + 0.577 \times 4 + 0.577 \times 16 \\ 0.408 \times 8 + 0.816 \times 4 + -0.408 \times 16 \end{bmatrix}}{11}$$

$$\theta = \begin{bmatrix} 6.9245 \\ 0 \end{bmatrix} \div 11$$

$$\theta = \begin{bmatrix} 0.62945 \\ 0 \end{bmatrix}$$

The following vector θ minimises the expression $\|x - B\theta\|^2 + \lambda\|\theta\|^2$ for the given matrix B and $\lambda = 10$.

Exercise 2 Vector Calculus Practices

Compute the following gradients over x or X.

a) $\dfrac{\partial x^T ABC x}{\partial x}$. $\qquad \dfrac{\partial x^T B x}{\partial x} = x^T (B + B^T)$ (identity)

$$\therefore \frac{\partial x^T ABC x}{\partial x} = x^T (ABC + (ABC)^T)$$

$$= x^T (ABC + C^T B^T A^T)$$

b) $\dfrac{\partial (Bx+b)^T C (Dx+d)}{\partial x}$

Product rule: $\dfrac{\partial fx}{\partial x} \cdot gx + fx \cdot \dfrac{\partial gx}{\partial x}$

$$\partial \underbrace{(Bx+b)^T}_{fx} \underbrace{C(Dx+d)}_{gx} / \partial x \Rightarrow \underbrace{C}_{fx} \cdot \underbrace{(Dx+d)}_{gx}$$

$$\frac{\partial (Bx+b)^T}{\partial x} \cdot C(Dx+d) + (Bx+b)^T \cdot \left(\frac{\partial C}{\partial x} \cdot Dx+d + C \cdot \frac{\partial Dx+d}{\partial x} \right)$$

$$= B^T \cdot C(Dx+d) + (Bx+b)^T \cdot (C \cdot D)$$

$$= B^T C D x + B^T C d + B^T x^T C D + b^T C D.$$

c) $\dfrac{\partial \operatorname{tr}(X^2)}{\partial X} = 2X.$

Exercise 3 concavity of a function

A function $f: \mathbb{R}^n \to \mathbb{R}$ with a convex domain is called a concave function iff its Hessian $H = \dfrac{\partial^2 f}{\partial x^2}$ is negative semidefinite. Consider:

$$f(x) = \left(\sum_{i=1}^{n} x_i^p \right)^{1/p}$$

with convex domain $\operatorname{dom}(f) = \mathbb{R}^n_{++}$ (n-dim strictly elementwise positive vectors), and $p < 1, p \neq 0$.

1. Evaluate the elementwise second order derivatives $\dfrac{\partial^2 f}{\partial x_i x_j}$ for arbitrary integer $i, j \in [1, n]$.

first order derivative:

$$\frac{\partial f x}{\partial x_i} = 1/p \left(\sum_{k=1}^{n} x_k^p \right)^{1/p - 1} \cdot p x_i^{p-1}$$

$$\frac{\partial f x}{\partial x_i x_j} = \cancel{p} x_i^{p-1} \cdot \frac{1}{\cancel{p}} \cdot \left(\frac{1}{p} - 1 \cdot \left(\sum_{k=1}^{n} x_k^p \right)^{\frac{1}{p} - 1 - 1} \right) \cdot \cancel{p} x_j^{p-1}$$

second order derivative

$$= (x_i x_j)^{p-1} \cdot \left(\frac{1}{p} - 1 \right) \left(\sum_{k=1}^{n} x_k^p \right)^{\frac{1}{p} - 2} \quad \text{for } i \neq j$$

2. Prove $H = (1-p) f(x)^{1-2p} \cdot (x^{p-1} \cdot x^{p-1^T} - f(x)^p \cdot \mathrm{diag}(x^{p-2}))$

$$\frac{\partial^2 f}{\partial x_i^2} = \frac{1}{p}\left(\frac{1}{p}-1\right)(f x)^{\frac{1}{p}-2} \cdot p x_i^{p-1} + \frac{1}{p}(fx)^{\frac{1}{p}-1} \cdot (p-1) p x_i^{p-2}$$

$$= \left(\frac{1}{p}-1\right) \cdot (f(x))^{\frac{1}{p}-2} \cdot x_i^{p-1} + (f(x))^{\frac{1}{p}-1} \cdot (p-1) x_i^{p-2}$$

The $(i,i)^{th}$ entry is $\frac{\partial^2 f}{\partial x^2}$ and the $(i,j)^{th}$ entry $(i \neq j)$ is $\frac{\partial^2 f}{\partial x_i x_j}$. $\frac{\partial^2 f}{\partial x_i^2}$ corresponds to the diagonal entries which are $-f(x)^p \, \mathrm{diag}(x^{p-2})$ in H.

$\frac{\partial^2 f}{\partial x_i x_j}$ corresponds to non-diag entries which are $x^{p-1} \cdot x^{p-1^T}$ in the equation H.

Both derivatives and their corresponding parts in the equation H seems to align.

So $H = (1-p) f(x)^{1-2p} \cdot (x^{p-1} x^{p-1^T} - f(x)^p \cdot \mathrm{diag}(x^{p-2}))$.

3. Prove H is negative semidefinite, hence f is concave.

To prove H is negative semidefinite, show that $\forall$ non-zero vectors v in $\mathbb{R}^n$ $v^T H v \leq 0$ is true.

Given $H = (1-p) f(x)^{1-2p} \cdot \left(x^{p-1} \cdot x^{p-1^T} - f(x)^p \cdot \mathrm{diag}(x^{p-2})\right)$

$$v^T H v = (1-p) f x^{1-2p} \left[v^T x^{p-1} x^{p-1^T} v - f(x)^p v^T \mathrm{diag}(x^{p-2}) v \right]$$

$v^T x^{p-1} x^{p-1^T} v$: This term is the square of the projection of z onto x^{p-1} so its non-negative.

$v^T \mathrm{diag}(x^{p-2}) v$: This term is the sum of the elementwise product of z and x^{p-2}. since x^{p-2} is elementwise positive ($x \in \mathbb{R}^n_{++}$), the sum is also non-negative.

And since $p < 1$, $f(x)^{1-2p}$ is positive. Thus,

$v^T x^{p-1} x^{p-1^T} v - f(x)^p v^T \mathrm{diag}(x^{p-2}) v$ is always non-negative.

$\therefore$ $v^T H v \leq 0$ and this shows that the Hessian is negative semidefinite, and given that the domain of f is convex, we can conclude that f is a concave function.

ercise 4 Expectations with respect to a Gaussian Distribution

common objective function in modern machine learning
the variational free-energy,

$$F(q(\theta)) = \int d\theta\, q\theta \log \frac{q(\theta)}{p\theta\, p(y|\theta,x)} = \int d\theta\, q\theta \left[\log q(\theta) - \log p(\theta) - \log p(y|\theta,x)\right]$$

onsider a simplified setting in which

$$p(\theta) = \mathcal{N}(\theta|0,1),$$
$$p(y|\theta,x) = \mathcal{N}(y|\theta x, \sigma_n^2),$$
$$q(\theta) = \mathcal{N}(\theta|\mu,\sigma^2),$$

where $\mathcal{N}(x;m,v)$ means x is a univariate Gaussian random variable with mean m and variance v.

1. Compute F.
2. Find the gradients $\frac{\partial}{\partial\mu}F$ and $\frac{\partial}{\partial\sigma}F$
3. Set these gradients to 0 and solve for μ and σ in terms of y, x and σ_n.

1. Compute F.

Integration rules:

$$\int \log(x)\,dx = x\log(x) - x + c \qquad \int e^x\,dx = e^x + c \qquad \frac{x^{n+1}}{n+1}$$

Gaussian distribution

$$N(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

$$\int q\theta\left[\log q\theta - \log p\theta - \log p(y|\theta,x)\right]d\theta.$$

Substitute $q(\theta)$, $p(\theta)$ and $p(y|\theta,x)$. And

For a standard Gaussian distribution, the total area under the curve of its pdf is always 1. $\therefore \int q(\theta) = 1$.

$$\int\left[\log\left(\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(\theta-\mu)^2}{2\sigma^2}\right)\right) - \log\left(\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{\theta^2}{2}\right)\right)\right.$$
$$\left. - \log\left(\frac{1}{\sqrt{2\pi\sigma_n^2}}\exp\left(-\frac{(y-\theta x)^2}{2\sigma_n^2}\right)\right)\right]d\theta$$

Using properties of logarithms to simplify:

$$\log(\exp(x)) = \exp(\log(x)) = x$$

$$x\log(a) = \log(a^x) \qquad \log(ab) = \log a + \log b$$

$$\int\left[\log\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) + \left(-\frac{(\theta-\mu)^2}{2\sigma^2}\right) - \log\left(\frac{1}{\sqrt{2\pi}}\right) - \left(-\frac{\theta^2}{2}\right)\right.$$
$$\left. - \log\left(\frac{1}{\sqrt{2\pi\sigma_n^2}}\right) - \left(-\frac{(y-\theta x)^2}{2\sigma_n^2}\right)\right]$$

$\longrightarrow$

$$\int \left[-\frac{1}{2}\log(2\pi\sigma^2) - \frac{(\theta-\mu)^2}{2\sigma^2} + \frac{1}{2}\log(2\pi) + \frac{\theta^2}{2} + \frac{1}{2}\log(2\pi\sigma_n^2) + \frac{(y-\theta x)^2}{2\sigma_n^2} \right] d\theta$$

$$\int -\frac{1}{2}\log(2\pi\sigma^2)\,d\theta - \int \frac{(\theta-\mu)^2}{2\sigma^2} + \int \frac{1}{2}\log(2\pi)\,d\theta + \int \frac{\theta^2}{2}\,d\theta + \int \frac{1}{2}\log(2\pi\sigma_n^2)\,d\theta + \int \frac{(y-\theta x)^2}{2\sigma_n^2}\,d\theta$$

$$= -\frac{1}{2}\log(2\pi\sigma^2)\theta - \frac{(\theta-\mu)^3}{6\sigma^2} + \frac{1}{2}\log(2\pi)\theta + \frac{\theta^3}{6} + \frac{1}{2}\log(2\pi\sigma_n^2)\theta + \frac{(y-\theta x)^3}{6\sigma_n^2} + C.$$

Rearrange & simplify using log properties

$$\left(\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(2\pi\sigma^2) + \frac{1}{2}\log(2\pi\sigma_n^2)\right)\theta - \frac{(\theta-\mu)^3}{6\sigma^2} + \frac{\theta^3}{6} + \frac{(y-\theta x)^3}{6\sigma_n^2} + C$$

$$= \frac{1}{2}\log(2\pi)\theta - \frac{(\theta-\mu)^3}{6\sigma^2} + \frac{\theta^3}{6} + \frac{(y-\theta x)^3}{6\sigma_n^2} + C$$

2. Find the gradients $\frac{\partial}{\partial \mu} F$ and $\frac{\partial}{\partial \sigma} F$

$$\frac{\partial F}{\partial \mu} = \frac{1}{2}\log(2\pi)\theta - \frac{(\theta-\mu)^3}{6\sigma^2} + \frac{\theta^3}{6} + \frac{(y-\theta x)^3}{6\sigma^2} + c$$

$$= 0 - \frac{3(\theta-\mu)^2 x - 1}{6\sigma^2} + 0 + 0 + 0$$

$$= (\theta-\mu)^2 / 2\sigma^2.$$

$$\frac{\partial F}{\partial \sigma} = \frac{1}{2}\log(2\pi)\theta - \frac{(\theta-\mu)^3}{6\sigma^2} + \frac{\theta^3}{6} + \frac{(y-\theta x)^3}{6\sigma^2} + c$$

$$= 0 + \frac{(\theta-\mu)^3}{3\sigma^3} + 0 - \frac{(y-\theta x)^3}{3\sigma^3} + 0$$

$$= \frac{(\theta-\mu)^3}{3\sigma^3} - \frac{(y-\theta x)^3}{3\sigma^3}$$

3. Set these gradients to zero & solve for μ & σ in terms of y, x, σ.

$$0 = \frac{(\theta-\mu)^2}{2\sigma^2} ; \sqrt{0} = \sqrt{(\theta-\mu)^2} ; 0 = \theta - \mu ; \theta = \mu$$

$$0 = \frac{(\theta-\mu)^3 - (y-\theta x)^3}{3\sigma^3} ; \sqrt[3]{0} = \sqrt[3]{(\theta-\mu)^3} - \sqrt[3]{(y-\theta x)}$$

$0 = (\theta-\mu) - (y-\theta x)$: if $\mu = \theta$ then $0 = -y + \theta x$;

$y = \theta x$; $\theta = \frac{y}{x}$ $\therefore \mu = \frac{y}{x}$ and $\sigma = \sigma$.