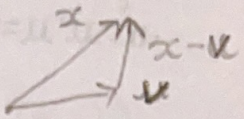
Assignment - 04

05/03/2025

1. (a) Let data $\{ \underbrace{x^{(1)}, x^{(2)} \dots, x^{(n)}}_{\text{zero mean and unit variance in each coordinate.}} \}$



define : $f_u(x) = \arg\min_{v \in \mathcal{V}} \| x - v \|^2 \qquad \mathcal{V} = \{ \alpha u : \alpha \in \mathbb{R} \}$

$= (x^T u)\, u$

Problem statement : proyet our data (d-dimensional) to k dimension (k=1 , here), we can formulate it as a minimisation problem, (the direction which gives best representation of our data

$$\min_{u : u^T u = 1} \sum_{i=1}^{n} \| x^{(i)} - f_u(x^i) \|_2^2$$

take,

$$\sum_{i=1}^{n} \| x^{(i)} - f_u(x^i) \|_2^2 = \sum_{i=1}^{n} \| x^{(i)} - ((x^i)^T u)\, u \|_2^2$$

$$= \sum_{i=1}^{n} (x^i - (x^{iT} u) u)^T (x^i - (x^{iT} u) u)$$

$$= \sum_{i=1}^{n} \{ (x^i)^T x^i - ((x^i)^T u)^2 - ((x^i)^T u) + (x^{iT} u)^2 u^T u \}$$

$$= \sum_{i=1}^{n} \{ (x^i)^T x^i - 2 ((x^i)^T u)^2 + (x^{iT} u)^2 u^T u \}$$

$$= \sum_{i=1}^{n} (x^i)^T x^i - (2 - u^T u)((x^i)^T u)^2$$

. simplified expression .

$$\arg\min_{u : u^T u = 1} \sum_{i=1}^{n} \| x^i - f_u(x^i) \|_2^2$$

$$\Leftrightarrow \arg\min_{u : u^T u = 1} \sum_{i=1}^{n} (x^i)^T x^i - (2 - u^T u)(x^{iT} u)^2$$

$$\Leftrightarrow \arg\max_{u : u^T u = 1} \sum_{i=1}^{n} ((x^i)^T u)^2 \quad (\because 2 - u^T u \geq 0 \text{ as } \|u\|^2 = 1)$$

$$\arg\max_{u:\,u^Tu=1} \sum_{i=1}^{n} (x^i)^T u)^2 \iff \arg\max_{u:\,u^Tu=1} \sum_{i=1}^{n} \{u^T x^i (x^i)^T u\}$$

$$\iff \arg\max_{u:\,u^Tu=1} u^T \left(\sum_{i=1}^{n} (x^i)(x^i)^T \right) u$$

but we know that,

$$\begin{bmatrix} \uparrow & \uparrow & & \uparrow \\ x^1 & x^2 & \dots & x^n \\ \downarrow & \downarrow & & \downarrow \end{bmatrix} \begin{bmatrix} \leftarrow & x^{1T} & \rightarrow \\ \leftarrow & x^{2T} & \rightarrow \\ & \vdots & \\ \leftarrow & x^{nT} & \rightarrow \end{bmatrix} = \sum_{i=1}^{n} (x^i)(x^i)^T$$

X^T $\quad$ X

$\therefore$

$$\iff \arg\max_{u:\,u^Tu=1} u^T (X^T X) u$$

Let $A = X^T X$ (symmentric and real)

$\therefore \exists\ Q \in \mathbb{R}^{d \times d},\ \Lambda \in \mathbb{R}^{d \times d}$ s.t

$A = Q \Lambda Q^T$ s.t Q = orthogonal matrix

Λ = diagonal matrix

$\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots \lambda_d)$

$\lambda_1 \geq \lambda_2 \geq \lambda \dots \lambda_d \rightarrow$ eigen vectors.

$\therefore$ we get

$$\iff \arg\max_{u:\,u^Tu=1} u^T (Q \Lambda Q^T) u$$

Λ_j → column / row of matrix Q.

For simplification,

$u^T (Q \Lambda Q^T) u$, let $u = \sum_{j=1}^{d} \alpha_j q_j$ (Q is invertible so forms basis for $\mathbb{R}^d$

$$= u^T (Q \Lambda Q^T) \sum_{j=1}^{d} \alpha_j q_j = u^T Q \Lambda \sum_{j=1}^{d} \alpha_j e_j$$

$$= u^T Q \sum_{j=1}^{d} \alpha_j \lambda_j e_j = u^T \sum_{j=1}^{d} \alpha_j \lambda_j q_j$$

$$= \sum_{j=1}^{d} \left\{ \alpha_j \lambda_j \left(q_j \sum_{i=1}^{d} \alpha_i q_i \right) \right\} = \sum_{j=1}^{d} \alpha_j \lambda_j \alpha_j q_j^T q_j$$

$$= \sum_{j=1}^{d} (\alpha_j)^2 \lambda_j$$

Now minimisation problem becomes (initial)

$$\arg\max \sum_{j=1}^{d} (\alpha_j)^2 \lambda_j \iff \arg\max \sum_{j=1}^{d} (\beta_j) \lambda_j$$

$$\|\alpha\|_2^2 = 1 \qquad \sum_{i=1}^{d} \beta_i = 1, \ \beta_i \geq 0 \qquad (\beta_j = \alpha_j^2)$$

$\beta = (\beta_1, \ldots \beta_d)$ maximising happens at one of the corner points of the region given by

$$\beta_{region} = \{ \beta : \sum_{i=1}^{n} \beta_i = 1, \ \beta_i \geq 0 \ \ i = 1 \ldots d \}$$

$\therefore$ the soln of the optimisation problem is

$\beta = (1, 0, 0 \ldots 0)$ i.e $\beta_1 = 1, \ \beta_2 \ldots \beta_d = 0$ $\because \lambda_1 \geq \lambda_k$ $(k = 2 \ldots d)$

$\iff \alpha = (1, 0, \ldots 0)$

$\iff u = v_1$ (eigen vector corresponding to λ_1)

1 (b) (i) $M \in \mathbb{R}^{p \times q}$

then $M^T M \in \mathbb{R}^{q \times q}$ ∴ Real matrix and square too

Now, $(M^T M)^T = M^T (M^T)^T = M^T M$ ∴ $M^T M$ is symmetric too

$M^T M = Q \Lambda Q^T$ $\Lambda = \text{diag}(\lambda_1, \lambda_2 \ldots \lambda_q)$

s.t $\lambda_1 \geq \lambda_2 \ldots \geq \lambda_q \in \mathbb{R}$

$Q \in \mathbb{R}^{q \times q}$ and orthogonal matrix

(ii) $M \in \mathbb{R}^{p \times q}$ (given)

$M = U \Sigma V^T$ where $U \in \mathbb{R}^{p \times p}$, $V \in \mathbb{R}^{q \times q}$.

where $U^T U = I$, $V^T V = I$

$\Sigma \in \mathbb{R}^{p \times q}$ and diagonal matrix.

$$M^T M = (U \Sigma V^T)^T (U \Sigma V^T)$$
$$= (V \Sigma^T U^T)(U \Sigma V^T)$$
$$= V \Sigma^T (U^T U) \Sigma V^T$$
$$= V \Sigma^T \Sigma V^T = V \Sigma^2 V^T \quad (\because \Sigma \text{ is diagonal matrix})$$

(iii)

q2 b)

original size (in bytes) = (512 × 512 × 3) × (8)
H W C
(512 × 512 × 3) = #pixels
(8) = size of each pixel (0-255) (in bytes)

compressed size (in bytes) = (512 × 512 × 1) × 4 + (16 × 3) × 8

(512 × 512 × 1): each pixel can be assigned a number between 0 to 15 depending on the centroid it is assigned to

4: size of each pixel (in bytes)

(16 × 3): 16 centroids having 3 color channel

8: each pixel's value range from (0-255)

$$\text{compressed factor} = \frac{\text{original size (bytes)}}{\text{compressed size (bytes)}}$$

$$\cong \frac{512 \times 512 \times 3 \times 8}{512 \times 512 \times 4 + 16 \times 3 \times 8}$$

$$\cong 6$$

q3 (a) T.P. $D_{KL}(P||Q) \geq 0 \;\forall P, Q$
and $D_{KL}(P||Q) = 0$ iff $P = Q$.

by Jensen's inequality,

$$E(f(x)) \geq f(E(x))$$

consider the function $f(x) = -\log x$, which is strictly convex. on using Jensen's inequality:

$$\sum_{x \in X} P(x) \log \frac{P(x)}{Q(x)} = E_P\left[\log \frac{P(x)}{Q(x)}\right] \geq \log E_P\left[\frac{P(x)}{Q(x)}\right]$$

$$= \log$$

$$D_{KL}(P||Q) = \sum_{x \in X} P(x) \log \frac{P(x)}{Q(x)}$$

$$E_P\left(-\log\left(\frac{Q(x)}{P(x)}\right)\right) \geq -\log E_P\left(\frac{Q(x)}{P(x)}\right)$$

expanding we get

$$-\sum_{x \in X} P(x) \log\left(\frac{Q(x)}{P(x)}\right) \geq -\log \sum_{x \in X}\left(P(x) \frac{Q(x)}{P(x)}\right)$$

$$D_{KL}(P||Q) \geq -\log 1 = 0 \quad \because \sum_{x \in X} Q(x) = 1$$

$$D_{KL}(P||Q) = \sum_{x \in X} P(x) \log \frac{P(x)}{Q(x)} \geq 0$$

equality holds iff argument of expectation fn satisfies $\log \frac{P(x)}{Q(x)} = 0 \;\forall x$ } for Jensen equality to hold.

$$\Leftrightarrow \frac{P(x)}{Q(x)} = 1 \quad \forall x \in X$$

$$\Leftrightarrow P(x) = Q(x) \quad \forall x \in X$$

q3 (b) Expanding joint KL divergence,

$$D_{KL}(P(x,y)\,\|\,Q(x,y)) = \sum_{x,y} P(x,y)\log\left(\frac{P(x,y)}{Q(x,y)}\right)$$

but $P(x,y) = P(x)\,P(y|x)$, $Q(x,y) = Q(x)\,Q(y|x)$

$$= \sum_{x}\sum_{y} P(x)\,P(y|x)\left[\log\left(\frac{P(x)}{Q(x)}\right) + \log\left(\frac{P(y|x)}{Q(y|x)}\right)\right]$$

$$= \sum_{x}\left[P(x)\log\frac{P(x)}{Q(x)}\left(\sum_{y} P(y|x)\right) + P(x)\sum_{y} P(y|x)\log\frac{P(y|x)}{Q(y|x)}\right]$$

$$= \sum_{x}\left[P(x)\log\frac{P(x)}{Q(x)}\sum_{y} P(y|x)\right] + \sum_{x}\left[P(x)\sum_{y} P(y|x)\log\frac{P(y|x)}{Q(y|x)}\right]$$

$$= D_{KL}(P(x)\,\|\,Q(x)) + E_{P(x)}\left[D_{KL}\,P(Y|X)\,\|\,Q(Y\|X)\right]$$

q3 (c) $D_{KL}(\hat{P}\,\|\,P_\theta) = \sum_{x} \hat{P}(x)\log\frac{\hat{P}(x)}{P_\theta(x)}$

$$= -\sum_{x} \hat{P}(x)\log P_\theta(x) + k \quad (k = \text{const.})$$

$$= -\sum_{x}\left\{\frac{1}{n}\sum_{i=1}^{n} 1\{x^i = x\}\right\}\log P_\theta(x) + k$$

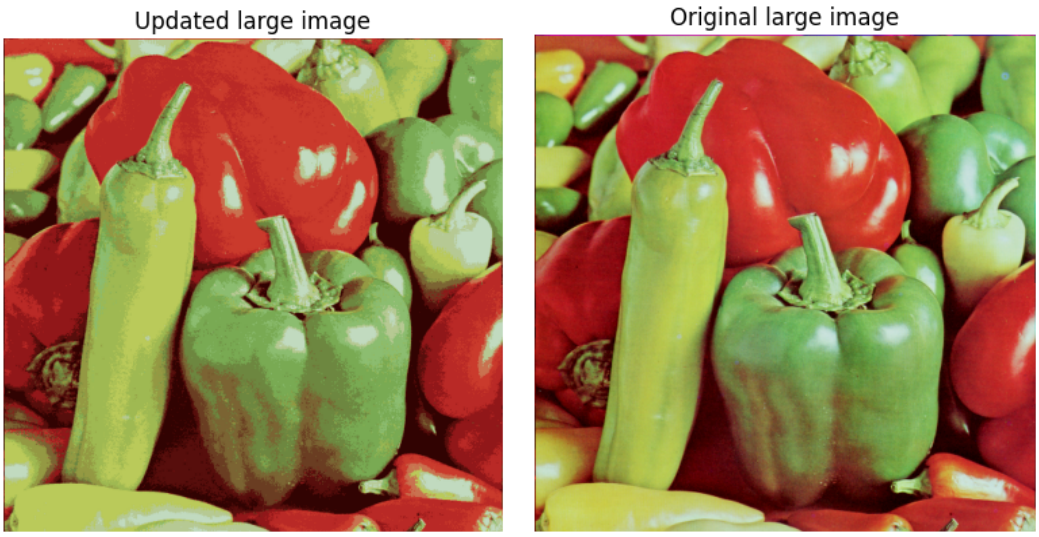
$$= -\sum_{i=1}^{n}\log P_\theta(x^i) + k$$

$\because \hat{P}(x) = \frac{1}{n}\sum_{i=1}^{n} 1\{x^i = x\}$

$\therefore$

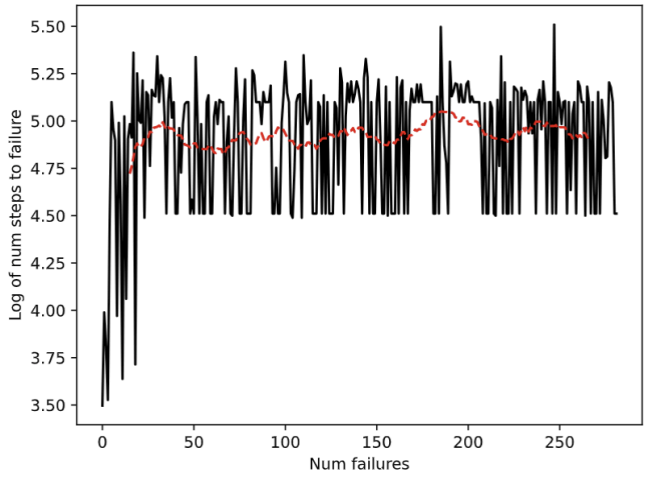
$$\operatorname{argmin}_\theta D_{KL}(\hat{P}\,\|\,P_\theta) \Leftrightarrow \operatorname{argmax}_\theta \sum_{i=1}^{n}\log P_\theta(x^i)$$

Q2 a)



Q2 b)

Q4 a)



Q4 b i)

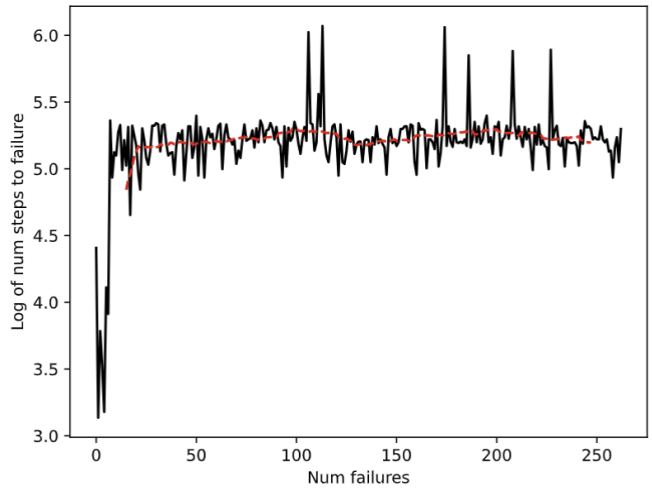
[INFO] Failure number 282

It took 282 trials for the algorithm to converge.

Q4 b ii)

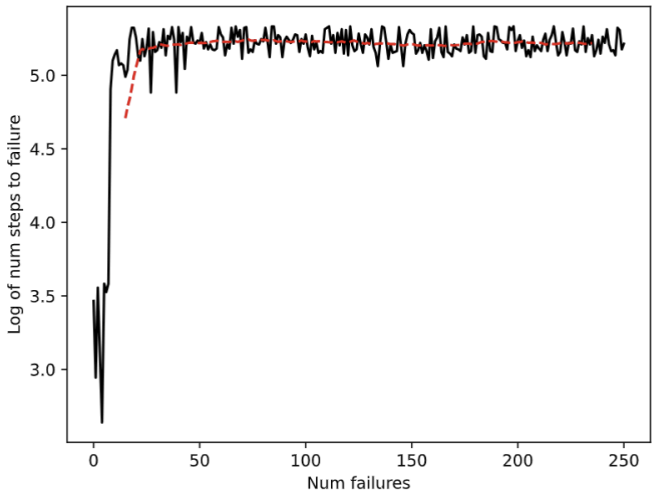
SEED = 1

[INFO] Failure number 263 (#trials before convergence)



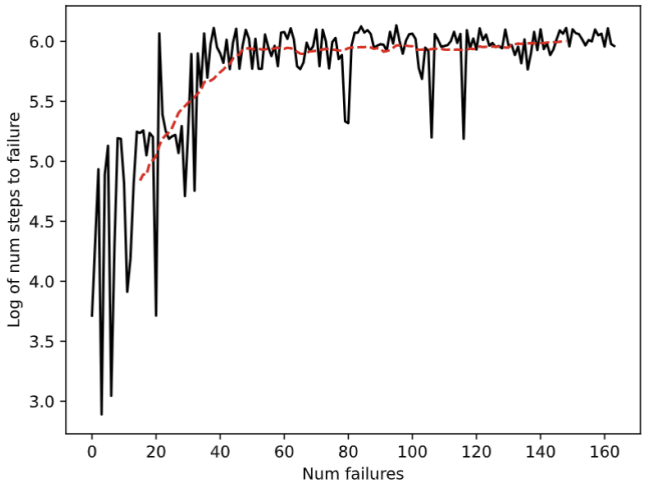
SEED = 2

[INFO] Failure number 251



SEED = 3

[INFO] Failure number 164



Different seeds cause different initialization and random transitions, therefore the algorithm's convergence changes