PSET-11 Q1)

→ Pick eigen vectors corresponding to largest eigen value of the covariance matrix Σ.

→ reduce to one dimension (PCA)

→ To reduce to two dimensions → pick ~~second largest vat~~ eigen vector corresponding to second largest eigen value.

→ We implement PCA in order to find best set of directions such that variability of data is maximized in the reduced dimensional space.

i.e $x' = Xp$

↓

new matrix after dimensional reduction. (PCA)

$\max \|x'\|^2 = \max \|Xp\|^2 \qquad p^T p = 1$

$= \max tr((Xp)^T (Xp)) \qquad$ such that $p^T p = 1$

$= \max (p^T X^T X p) \qquad$ such that $p^T p = 1$

$= \max (p^T S p) \qquad$ such that $p^T p = 1$

where p is the direction.

S = Scatter matrix (Un normalized Covariance matrix)

Using Lagrangian multipliers:-

$\max (p^T S p) \quad s.T \; p^T p = 1$

$\rightarrow \max (p^T S p) - \lambda (p^T p - 1) \quad (\lambda > 0)$

This is used to solve constrained optimization by introducing λ to combine $p^T S p$ and the constraint

Then solve for stationary point:

$$\frac{\partial L(P,\lambda)}{\partial \lambda} = 0 \qquad \text{where } L(P,\lambda) = P^T S P - \lambda(P^T P - 1)$$

$$\rightarrow P^T P = 1$$

$$\frac{\partial L}{\partial P} \rightarrow SP = \lambda P$$

$$\therefore L(P,\lambda) = P^T \lambda P - \lambda(1-1) = \lambda P^T P = \lambda$$

$\therefore$ optimization reduces to max λ such that $\lambda P = SP$

This mean eigen vector ~~of~~ P of S is found corresponding to maximum eigenvalue λ.

To find K directions, find eigen vector corresponding to each of K eigenvalues sorted in descending order.

i.e take P_1 & P_2 which $SP = \lambda P$

P_1 is the eigen vector corresponding to highest λ.

$$\lambda = \begin{bmatrix} \lambda_1 & 0 & \cdots & \\ 0 & \lambda_2 & \cdots & \\ 0 & 0 & \lambda_3 & \cdots \\ 0 & 0 & 0 & \cdots \lambda_n \end{bmatrix}$$

$\therefore$ To maximize value, P_2 should correspond to second highest eigen vector.

② Let $x_i \in \mathbb{R}^n$ be a sample

& let X be the matrix of all samples x_i.

$\Rightarrow X \in \mathbb{R}^{m \times n}$

Dimensionality reduction using PCA:-

On applying PCA to X, we obtain a matrix of principle components, $V \in \mathbb{R}^{n \times d}$ (assuming $d = \text{rank}(X)$)

→ Reconstructed matrix $\boxed{X' = XVV^T}$

Loss function $J = \frac{1}{m} \sum_{i=1}^{m} (\text{original matrix} - \text{reconstructed matrix})^2 \quad \forall x_i \in X$

$$= \frac{1}{m} \sum_{i=1}^{m} (x_i' - x_i)(x_i' - x_i)^T$$

$$= \frac{1}{m} \sum_{i=1}^{m} (x_i VV^T - x_i)(x_i VV^T - x_i)^T$$

Minimising the objective function J using gradient descent:-

J is a function of V $\quad (\because X' = XVV^T)$

$$\theta_j \leftarrow \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_j)$$

Here $\theta_j \equiv V$

$$V \leftarrow V - \alpha \frac{\partial}{\partial V} \left[\frac{1}{m} \sum_{i=1}^{m} (x_i VV^T - x_i)(x_i VV^T - x_i)^T \right]$$

$$\frac{dJ(V)}{dV} = \frac{d}{dV} \left(\frac{1}{m} \sum_{i=1}^{m} (x_i VV^T - x_i)(x_i VV^T - x_i)^T \right)$$

$$= \frac{1}{m} \sum_{j=1}^{m} \left[2(x_i VV^T - x_i)^T (2x_i V) \right]$$

$$= \frac{4}{m} \sum_{i=1}^{m} \left[x_i (VV^T - I) \right]^T (x_i V) = \frac{4}{m} \sum_{i=1}^{m} (VV^T - I)(x_i^T x) V$$

Scanned by CamScanner

Hence, update function

$$V \leftarrow V - \frac{4\alpha}{m} \sum_{i=1}^{m} (VV^T - I)(x_i^T x)V$$

Hence, update function

$$V \leftarrow V - \frac{4\alpha}{m} \sum_{i=1}^{m} (VV^T - I)(x_i^T x) V$$

3. In case of PCA

Loss function = reconstruction error.

each sample $x_i \in \mathbb{R}^n$

Given sample data $X \in \mathbb{R}^{m \times n}$

Reconstructed data $X' = XVV^T$

where $V \in \mathbb{R}^{r \times d}$ is matrix of principal components.

$$RSS = J = \frac{1}{m} \sum (\text{error})^2 \quad \forall \text{ samples.}$$

$$= \frac{1}{m} \sum_{i=1}^{m} (x_i' - x_i)(x_i' - x_i)^T \quad \forall x_i \in X$$

$$\Rightarrow J(V) = \frac{1}{m} \sum_{i=1}^{m} (x_i VV^T - x_i)(x_i VV^T - x_i)^T \quad \forall x_i \in X$$

$$\frac{dJ(V)}{dV} = \frac{4}{m} \sum_{i=1}^{m} (VV^T - I)(x_i^T x) V$$

(i) For L_1 regularisation (Lasso regression) $\Rightarrow$ RSS + L_1-norm

$$L^{lasso}(\lambda) = (J(V) + \lambda \|V\|_1)$$

$$\Rightarrow \min \left(\frac{1}{m} \sum_{i=1}^{m} [x_i VV^T - x_i][x_i VV^T - x_i]^T + \lambda \|V\|_1 \right)$$

$$\Rightarrow \frac{\partial}{\partial V} L^{lasso}(\lambda) = 0$$

$$\Rightarrow \frac{4}{m} \sum_{i=1}^{m} (VV^T - I)(x_i x_i^T) V + \lambda (\text{sign}(V)) = 0$$

(ii) For L_2 regularisation (Ridge regression) $\rightarrow$ RSS + L2-norm.

$$L^{Ridge}(\lambda) = (J(\cancel{W}) + \lambda \|V\|_2)$$

$$\Rightarrow \min \left(\frac{1}{m} \sum_{i=1}^{m} [x_i V V^T - x_i] [x_i V V^T - x_i]^T + \lambda \|V\|_2 \right)$$

$$\Rightarrow \frac{\partial}{\partial V} L^{Ridge}(\lambda) = 0$$

$$\Rightarrow \frac{4}{m} \sum_{i=1}^{m} (VV^T - I)(x_i^T x_i) V + \lambda \frac{V}{\|V\|_2} = 0$$