02/28/2023 Generative Models (Probablistic PCA) Continued

$P(z) = N(0, I)$ ← assumption

$P(x|z) = N(Wz + \mu, \sigma^2 I)$

$P(x) = \int P(x|z)P(z)dz = N(\mu, \underbrace{\sigma^2 I + WW^T}_{\Psi})$ → maximize the marginal log-likelihood

- optimization problem (max log-likelihood Problem)

$$\max_{\theta=[W,\mu,\sigma^2]} \sum_{i=1}^{n} \log P_\theta(x_i)$$

$$= \max_{\theta=[W,\mu,\sigma^2]} \left\{ \frac{-nd}{2}\log(2\pi) - \frac{n}{2}\log(|\Psi|) - \frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^T\Psi^{-1}(x_i-\mu) \right\}$$

$W \in \mathbb{R}^{d \times m}$ $\quad z \in \mathbb{R}^m$ $\quad x \in \mathbb{R}^d$

- maximum log-likelihood estimates (ML)

$$\hat{\mu}_{ML} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\hat{W}_{ML} = U_m(L_m - \sigma^2 I)^{1/2} R$$

$$\hat{\sigma}^2_{ML} = \frac{1}{d-m}\sum_{i=m+1}^{d} \lambda_i$$

$R$: some orthogonal matrix (square matrix $R^T R = I = RR^T$)

$U_m$: $d \times m$ matrix whose columns are eigenvectors of $S$

$L_m$: $m \times m$ diagonal matrix whose elements are eigenvalues of $S$

$S$: Covariance matrix $S = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x})^T$

- if we solve max log-likelihood problem using SVD/PCA, then $R = I$

- Computing max log-likelihoods through SVD

1) $X = \begin{bmatrix} — x_1 — \\ — x_2 — \\ \vdots \\ — x_n — \end{bmatrix}_{n \times d}$

2) $A = \begin{bmatrix} — x_1 - \bar{x} — \\ — x_2 - \bar{x} — \\ \vdots \\ — x_n - \bar{x} — \end{bmatrix}_{n \times d}$

3) Do SVD of A

$A = U \Sigma V^T$ $\quad (S = \frac{1}{n} A^T A)$

4) Compute $U_m = V[:, :m]$

$L_m = \Sigma[:m, :m]$

5) Compute $\hat{W}_{ML}, \hat{\sigma}^2_{ML}, \hat{\mu}_{ML}$

$\hat{\mu}_{ML} = \frac{1}{n} \sum_{i=1}^{n} x_i$

$\hat{\sigma}^2_{ML} = \frac{1}{d-m} \sum_{i=m+1}^{d} \lambda_i = \frac{1}{d-m} \sum_{i=m+1}^{d} \sigma_i^2$

$\hat{W}_{ML} = U_m (L_m - \hat{\sigma}^2_{ML} I)^{1/2}$

$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}$ example

$A^T = \begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{bmatrix}$

$= \frac{1}{6} \underset{2 \times 3}{\begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{bmatrix}} \underset{3 \times 2}{\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}}$

$= \frac{1}{6} \begin{bmatrix} 1+9+25 & 2+12+30 \\ 2+12+30 & 4+16+36 \end{bmatrix}$

- SVD & PCA are very expensive computationally ∴ we use an iterative method with LARGE datasets