Generative Models (Probablistic PCA) Continued 02/23/2023

- probability review (marginal)

		x: 1	2	
y	3	0.1	0.3	0.4
	4	0.3	0.3	0.6
		0.4	0.6	1

$P(x=1) = P(x=1 \mid y=3)\, P(y=3) + P(x=1 \mid y=4)\, P(y=4)$

$= P(x=1, y=3) + P(x=1, y=4)$

$= 0.1 + 0.3$

$= 0.4$

- types
  - joint $P(x,y)$
  - marginal $P(x)$ or $P(y)$
  - conditional $P(x \mid y)$ or $P(y \mid x)$

$$P(x \mid y) = \frac{P(x,y)}{P(y)}$$

- probablistic PCA
  - for simplified model $\theta = \phi$

$P_\lambda(z) = N(0, I)$

$P_\theta(x \mid z) = N(Wz + \mu, \sigma^2 I)$

$P_\theta(x) = ?$

$q_\phi(z \mid x) = ?$ [encoder]

$P(z) = N(\alpha, \Lambda^{-1})$

$P(x \mid z) = N(Az + b, L^{-1})$

$P(x) = N(A\alpha + b, L^{-1} + A\Lambda^{-1}A^T)$

$P(z \mid x) = N(\Sigma \{A^T L(z-b) + \Lambda\alpha\}, \Sigma); \Sigma = (\Lambda + A^T L A)^{-1}$

$$= \frac{P(x \mid z) P(z)}{P(x)}$$

$W = A$ $\quad \mu = b$ $\quad I = \Lambda^{-1}$

$z = z$ $\quad \sigma^2 I = L^{-1}$ $\quad \theta = \alpha$

$P_\theta(x) = N(\mu, \sigma^2 I + WW^T)$

$\Sigma = (I + W^T(\frac{1}{\sigma^2} I) W)^{-1} = \sigma^2(\sigma^2 I + W^T W)^{-1}$

$q_\phi(z|x) = N(\sigma^2(\sigma^2 I + W^T W)^{-1}\{W^T(\frac{1}{\sigma^2} I)(z-\mu),$
$\sigma^2(\sigma^2 I + W^T W)^{-1})$

$= N((\sigma^2 I + W^T W)^{-1} W^T(z-\mu), \sigma^2 \underbrace{(\sigma^2 I + W^T W)^{-1}}_{M})$

$= N(M^{-1} W^T(z-\mu), \sigma^2 M^{-1})$

- how do we maximize the likelihood?
  - we need to find W, μ, and σ to generate new samples
  - since this is a result of a linear model ($x = Wz + \mu + \epsilon$), this is a linear model
  1.) assume we have n samples that are independentally and identically distributed (i.i.d.)

$D = \{x_i\}_{i=1}^n = \{x_1, x_2, \dots x_n\}$

$\max_\theta P_\theta(D) = \prod_i P_\theta(x_i)$

- these are independent because the samples are independent $[P(A \cap B) = P(A|B)P(B) = P(A)P(B)$!

$$\max_\theta \log\left(\prod_{i=1}^{n} P_\theta(x_i)\right) = \max_\theta \sum_{i=1}^{n} \log P_\theta(x_i)$$

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

$$\max_\theta \sum_{i=1}^{d} \log\left(\frac{1}{2\pi^{d/2}|\Psi|^{1/2}} e^{-\frac{1}{2}(x_i-\mu)^T\Psi^{-1}(x_i-\mu)}\right)$$

determinant not abs value!

$$= -\frac{nd}{2}\log(2\pi) - \frac{n}{2}\log|\Psi| - \frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^T\Psi^{-1}(x_i-\mu)$$

where $\Psi = \sigma^2 I + WW^T$

- Ψ is a positive definite ($\sigma > 0$) matrix meaning the eigenvalues are positive and the determinant is positive
- $x^T\Psi x > 0 \;\forall x$
- maximum log-likelihood estimates (ML)

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

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

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