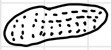
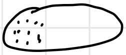
(1) 是否足以表征真实分布



(2) 是否不足以表征真实分布 manifold

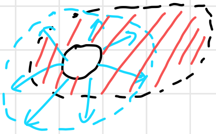


observed distribution o

True distribution

underlying distribution ///

Augmented observed distribution



information augmentation

EIG (effective information gain) trade-off $\begin{cases} \text{richness} \\ \text{bias} \end{cases}$

$X \in R^{d \times N}$ d dimension, N samples

covariance matrix

$$\Sigma = E[\frac{1}{N}\sum_{i=1}^{N} x_i x_i^T] = \frac{1}{N} XX^T \in R^{d \times d}$$

manifold volume as:

$\sqrt{\det(\frac{1}{N}XX^T)}$ ⇒ Singular value decomposition (SVD)

$(\frac{1}{N}XX^T)$ not full rank ⇒ lead volume 0

⇒ $\sqrt{\det(\frac{1}{N}XX^T + I)}$ (positive definite matrix) (without affect monotonicity)

numerical stability

$V(X) = \frac{1}{2}\log_2 \det(I + \frac{1}{N}XX^T)$ ⇒ Richness (measure)

constraint: (derive EIG)

① $Z' = 0$ EIG = 0

② $Z' \neq 0$, but focus on a point

information → 0 cost ↑

EIG < 0 → define lower bound

Definition:

$X \xrightarrow{A} X + X'$ (samples, augmented samples)

$N + N'$ (samples number)

$X + X' \xrightarrow{\text{Train}} M$

Feature Extract

$Z, Z' \longrightarrow F = [Z, Z'] \in R^{d \times (N+N')}$

normalize

$EIG(A, F, Z) = \frac{V(F) - V(Z)}{V(Z)}$ → $\begin{cases} \text{(1) In image space richness constant} \\ \text{(2) In embedding space richness} \\ \quad \propto \text{ parameter initial random} \\ \quad\quad \downarrow \\ \quad \text{differ from A} \end{cases}$

So we focus ration, rather that real value. intrinsic change ratio. (information augmentation)

① constraint proved:

$$EIG(A, F, Z) = \frac{V(F) - V(Z)}{V(Z)}$$

$$= \frac{\frac{1}{2}\log_2 \det(I + \frac{1}{N+N'}FF^T) - \frac{1}{2}\log_2 \det(I + \frac{1}{N}ZZ^T)}{\frac{1}{2}\log_2 \det(I + \frac{1}{N}ZZ^T)}$$

$$= \frac{\log_2 \frac{\det(I + \frac{1}{N+N'}FF^T)}{\det(I + \frac{1}{N}ZZ^T)}}{\log_2 \det(I + \frac{1}{N}ZZ^T)} \quad (\log_a b \Leftrightarrow \frac{\ln b}{\ln a})$$

$$= \log_S \frac{\det(I + \frac{1}{N+N'}FF^T)}{\det(I + \frac{1}{N}ZZ^T)} \quad (S: \det(I + \frac{1}{N}ZZ^T))$$

$Z' = 0$ ?⇒ EIG = 0

$EIG = \log_S 1 = 0$ $Z' = 0$ ⇒ EIG = 0 prove complete!

② constraint proved: lower bound → $\frac{-N'}{N+N'}$

$\begin{rcases} \log x: \text{concave function} \to \\ \det(\cdot): \text{Matrix volume} \\ S: \text{positive definite matrix} \\ \frac{\partial \log\det(S)}{\partial S} = S^{-1} \Rightarrow \log\det(S) \text{ is concave} \end{rcases}$

concave function: Jensen unequation

$f(\sum_{j=1}^{k} \alpha_j x_j) \geq \sum_{j=1}^{k} \alpha_j f(x_j)$, $\alpha_j > 0$ $\sum_{j=1}^{k} \alpha_j = 1$

↳ $\log\det(\sum_{j=1}^{k} \alpha_j S_j) \geq \sum_{j=1}^{k} \alpha_j \log\det(S_j)$

So we have $\log\det(\sum_{j=1}^{k} \alpha_j S_j) \geq \sum_{j=1}^{k} \alpha_j \log\det(S_j)$

$(\alpha_j > 0, \sum_{j=1}^{k} \alpha_j = 1, S_j (1, 2, \cdots k))$

Assume $k = 2$:

$$\log\det(\alpha_1 S_1 + \alpha_2 S_2) \geq \alpha_1 \log\det(S_1) + \alpha_2 \log\det(S_2)$$

$\alpha_1 = \frac{N}{N+N'}$ $\alpha_2 = \frac{N'}{N+N'}$ $S_1 = I + \frac{1}{N}ZZ^T$ $S_2 = I + \frac{1}{N'}Z'Z'^T$

$$\Rightarrow \log\det(\alpha_1 S_1 + \alpha_2 S_2) = \log\det(I + \frac{1}{N+N'}(ZZ^T + Z'Z'^T))$$

$$= \log\det(I + \frac{1}{N+N'}(FF^T))$$

Then we have:

$$\log\det\left(I+\frac{1}{N+N'}FF^T\right) \geq \frac{N}{N+N'}\log\det\left(I+\frac{1}{N}ZZ^T\right)+\frac{N'}{N+N'}\log\det\left(I+\frac{1}{N'}Z'Z'^T\right)$$

$$\frac{1}{2}\log\det\left(I+\frac{1}{N+N'}FF^T\right) \geq \frac{1}{N+N'}\left[\frac{N}{2}\log\det\left(I+\frac{1}{N}ZZ^T\right)+\frac{N'}{2}\log\det\left(I+\frac{1}{N'}Z'Z'^T\right)\right]$$

$$V(F) \geq \frac{1}{N+N'}\left(NV(Z)+N'V(Z')\right)$$

$$EIG(A,F,Z) = \frac{V(F)-V(Z)}{V(Z)} \geq \frac{\left[\frac{N}{N+N'}V(Z)+\frac{N'}{N+N'}V(Z')\right]-V(Z)}{V(Z)}$$

$$\Rightarrow EIG \geq \frac{\left[\left(\frac{N}{N+N'}-\frac{N+N'}{N+N'}\right)V(Z)+\frac{N'}{N+N'}V(Z')\right]}{V(Z)}$$

$$\Rightarrow EIG \geq \frac{\frac{-N'}{N+N'}V(Z)+\frac{N'}{N+N'}V(Z')}{V(Z)}$$

$(I+\frac{1}{N'}Z'Z'^T)$ is position definite matrix, $V(Z')\geq 0$

when $Z'$ focus on a point:

$V(Z')\to 0$ So $EIG \to$ lower bound $\frac{-N'}{N+N'}$

prove completed!