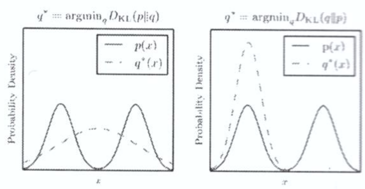
Home Work Assignment 1
Due date: Nov. 14th

1. Given a vector of random variables $\mathbf{x} \in \mathbb{R}^n$ with covariance matrix Σ, $\Sigma_{ij} = \mathbb{E}[\mathbf{x}_i\mathbf{x}_j] - \mathbb{E}[\mathbf{x}_i]\mathbb{E}[\mathbf{x}_i]$.
   (a) If $\mathbf{w} \in \mathbb{R}^n$ and $y = \mathbf{w}^T\mathbf{x} = \sum_i^n \mathbf{w}_i\mathbf{x}_i$, show that $\mathbf{w}^T\Sigma\mathbf{w} = \text{var}(y)$.
   (b) Show that Σ is positive semi-definite.

2. We define the KL divergence as $KL(p||q) = \mathbb{E}_{x\sim p}\left[\log(\frac{p(x)}{q(x)})\right] = \mathbb{E}_{x\sim p}\left[-\log(\frac{q(x)}{p(x)})\right]$ (infinity if q(x)=0 and $p(x) > 0$).
   (a) Prove that $KL(p||q) \geq 0$ and $KL(p||q) = 0$ if and only if $p = q$ (hint: Jensen inequality)
   (b) Show that minimizing the KL divergence $\theta^* = \arg\min_\theta KL(p||q_\theta)$ (note that it isn't symmetric so order matters) is the same as maximum likelihood $\theta^* = \arg\max_\theta \mathbb{E}_p[q_\theta(x)]$
   Note: we usually minimize negative log-likelihood instead, and when we approximate with samples $x_1, ..., x_N \sim p$ we get $\theta^* = \arg\min_\theta \frac{1}{N}\sum_i -\log(q_\theta(x_i))$
   (c) Show that if $p(x_1, ..., x_n) = \prod p_i(x_i)$ and $q(x_1, ..., x_n) = \prod q_i(x_i)$ are independent then $KL(p||q) = \sum_i KL(p_i||q_i)$.
   (d) Compute the KL divergence between two Gaussians $p(x) \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $q(x) \sim \mathcal{N}(\mu_2, \sigma_2^2)$.

3. In the figure you can see what happens when we try to fit a double Gaussian (distribution p) with a single Gaussian (q) using both $KL(p||q)$ and $KL(q||p)$. One can see that $q^* = \arg\min_q KL(p||q)$ is inclusive (tries to cover the entire support of p) while $q^* = \arg\min_q KL(q||p)$ is exclusive (tries to not go beyond the support). Try to explain this behavior from the KL divergence definition.



מודלים גנרטיביים עמוקים

תרגיל #1

(1) $x \in \mathbb{R}^n \qquad cov(x) = \Sigma$

$$\Sigma_{ij} = E(x_i x_j) - E(x_i)E(x_j)$$

(a) $w \in \mathbb{R}^n \qquad y = w^T x = \sum_i^n w_i x_i$

$$\begin{aligned}
Var(y) &\triangleq E(y^2) - E^2(y) \\
&= E(w^T x x^T w) - E(w^T x)\left[E(w^T x)\right]^T \\
&= w^T E(x x^T) - w^T E(x) E^T(x) w \\
&= w^T \underbrace{\left[E(x x^T) - E(x) E^T(x)\right]}_{= Cov(x) = \Sigma} w
\end{aligned}$$

$$\Downarrow$$

$$Var(y) = w^T Cov(x) w$$

(b) $Cov(x) = E\left\{[x - E(x)][x - E(x)]^T\right\}$

$$\Downarrow$$

$$\begin{aligned}
w^T Cov(x) w &= w^T E\left\{[x - E(x)][x - E(x)]^T\right\} w \\
&= E\left\{\underbrace{w^T [x - E(x)]}_{= s}[x - E(x)]^T w\right\} \\
&= E\{s^2\}
\end{aligned}$$

הערך הריבועי של s מספר ממשי תהיה תמיד גדולה-שווה ל-0.

ולכן התוחלת של s^2 תהיה גדולה-שווה ל-0 גם כן:

$$s^2 \geq 0 \Rightarrow E(s^2) \geq 0 \Rightarrow w^T Cov(x) w \geq 0$$

(2) $$KL(p\|q) = E_{x\sim p}\left\{\log\left[\frac{p(x)}{q(x)}\right]\right\} = E_{x\sim q}\left\{-\log\left[\frac{q(x)}{p(x)}\right]\right\}$$

(a) $$KL(p\|q) = E_{x\sim p}\left\{-\log\left[\frac{q(x)}{p(x)}\right]\right\}$$

עפ"י אי-שוויון ינסן:

עבור פונ' f קונווקסית $f, g : \mathbb{R} \to \mathbb{R}$

$$E\left[f(g(x))\right] \geq f\left(E\left[g(x)\right]\right)$$

נקח: $f(x) = -\log(x)$ $g(x) = \frac{p(x)}{q(x)}$

ולכן:

$$E_{x\sim p}\left\{-\log\left[\frac{p(x)}{q(x)}\right]\right\} \geq -\log\left\{E_{x\sim p}\left[\frac{p(x)}{q(x)}\right]\right\}$$

עבור $p=q$ מתקבל:

$$KL(p\|q) \geq -\log\left\{E_{x\sim p}\left[\frac{p(x)}{q(x)}\right]\right\} = -\log\left[E(1)\right] =$$
$$= -\log(1) = 0$$

ולכן:

$KL(p\|q) \geq 0$ אם ורק אם $p=q$

(b) $$\theta^* = \arg\min_\theta KL(p \| q_\theta)$$

עפ"י הגדרה:

$$KL(p \| q_\theta) = E_{x\sim p}\left\{\log\left[\frac{p(x)}{q_\theta(x)}\right]\right\}$$

$$\theta^* = \arg\min_\theta E_{x\sim p}\left\{\log\left[\frac{p(x)}{q_\theta(x)}\right]\right\}$$

$$= \arg\min_\theta E_{x\sim p}\left\{\log[p(x)] - \log[q_\theta(x)]\right\}$$

פונקצית ההתפלגות $p(x)$ אינה משפיעה על בעית האופטימיזציה של הפרמטרים θ ולכן ניתן להסיר את האיבר הראשון:

$$\theta^* = \arg\min_\theta E_{x\sim p}\left\{-\log[q_\theta(x)]\right\}$$

$$= \arg\min_\theta -E_{x\sim p}\left\{\log[q_\theta(x)]\right\}$$

מציאת המינימום של $-E(\log(q(x)))$ שווה למציאת המקסימום של $E(\log(q(x)))$ ולכן מתקבל:

$$\theta^* = \arg\min_\theta E_{x\sim p}\left\{\log\left[\frac{p(x)}{q_\theta(x)}\right]\right\} = \arg\max_\theta E_{x\sim p}\left\{\log[q_\theta(x)]\right\}$$

(c)

נתון כי המשתנים p ו- q בלתי תלויים

$$p(x_1,\dots,x_n) = \prod p_i(x_i)$$
$$q(x_1,\dots,x_n) = \prod q(x_i)$$

$$KL(p \| q) = E_{x\sim p}\left\{\log\left[\frac{p(x)}{q(x)}\right]\right\}$$

$$= E_{x\sim p}\left\{\log[p(x)] - \log[q(x)]\right\}$$

$$= E_{x\sim p}\left\{\log\left[\prod p_i(x_i)\right] - \log\left[\prod q_i(x_i)\right]\right\}$$

$$= E_{x\sim p}\left\{\sum_i \log(p_i(x_i)) - \sum_i \log(q_i(x_i))\right\}$$

$$= E_{x\sim p}\left\{\sum_i \log\left[\frac{p_i(x_i)}{q_i(x_i)}\right]\right\} = \sum_i E_{x\sim p}\left\{\log\left[\frac{p_i(x_i)}{q_i(x_i)}\right]\right\}$$

$$= \sum_i KL(p_i(x_i) \| q_i(x_i))$$

(d) $p(x) \sim N(\mu_1, \sigma_1^2)$

$q(x) \sim N(\mu_2, \sigma_2^2)$

$$KL(p\|q) = E_{x\sim p}\left\{\log\left[\frac{p(x)}{q(x)}\right]\right\} = E_{x\sim p}\left\{\log[p(x)] - \log[q(x)]\right\}$$

$$p(x) = \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left[\frac{-(x-\mu_1)^2}{2\sigma_1^2}\right]$$ [illegible]:

$$q(x) = \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp\left[-\frac{(x-\mu_2)^2}{2\sigma_2^2}\right]$$

$$= E_{x\sim p}\left\{\cancel{-\frac{1}{2}\log(2\pi)} - \log(\sigma_1) - \frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2 + \cancel{\frac{1}{2}\log(2\pi)} + \log(\sigma_2) + \frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2\right\}$$

$$= E_{x\sim p}\left\{\log\left(\frac{\sigma_2}{\sigma_1}\right) + \frac{1}{2}\left[\left(\frac{x-\mu_2}{\sigma_2}\right)^2 - \left(\frac{x-\mu_1}{\sigma_2}\right)^2\right]\right\}$$

$$= \log\left(\frac{\sigma_2}{\sigma_1}\right) + \frac{1}{2\sigma_2^2}E_{x\sim p}\left\{(x-\mu_2)^2\right\} - \frac{1}{2\sigma_1^2}\underbrace{E_{x\sim p}\left\{(x-\mu_1)^2\right\}}_{=\sigma_1^2}$$

$$= \log\left(\frac{\sigma_2}{\sigma_1}\right) + \frac{1}{2\sigma_2^2}E_{x\sim p}\left\{(x-\mu_2)^2\right\} - \frac{1}{2\cancel{\sigma_1^2}}\cdot\cancel{\sigma_1^2}$$

$$= \log\left(\frac{\sigma_2}{\sigma_1}\right) + \frac{1}{2\sigma_2^2}E_{x\sim p}\left\{(x-\mu_2)^2\right\} - \frac{1}{2}$$

נחליף את הביטוי $(x-\mu_2)^2$ בביטוי הבא:

$$(x-\mu_2)^2 = (x-\mu_1+\mu_1-\mu_2)^2 =$$

$$= (x-\mu_1)^2 + 2(x-\mu_1)(\mu_1-\mu_2) + (\mu_2-\mu_2)^2$$

בהצבה מתקבל:

$$= \log\left(\frac{\sigma_2}{\sigma_1}\right) + \frac{1}{2\sigma_1^2}\left\{\underbrace{E_{x\sim p}\left[(x-\mu_1)^2\right]}_{=\sigma_1^2} + 2(\mu_1-\mu_2)\underbrace{E_{x\sim p}\left[(x-\mu_1)\right]}_{=0}\right.$$

$$\left. + (\mu_1-\mu_2)^2\right\} - \frac{1}{2}$$

ולכן מתקבל:

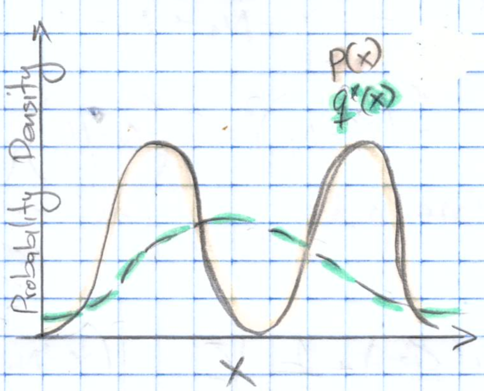
$$KL(p\|q) = \log\left(\frac{\sigma_2}{\sigma_1}\right) + \frac{\sigma_1^2 + (\mu_1-\mu_2)^2}{2\sigma_2^2} - \frac{1}{2}$$

3

עבור המקרה השמאלי:

(a) $q^* = \underset{q}{\arg\min}\, KL(p\|q)$

$= \underset{q}{\arg\min}\left\{E_{x\sim p}\left[\log\left(\frac{p(x)}{q^*(x)}\right)\right]\right\}$



אנו שואפים למזער את הערך: $\log\left(\frac{p(x)}{q^*(x)}\right)$. הערך המינימלי יתקבל כאשר היחס בין $p(x)$ ל-$q(x)$ יתקרב ל-1, כלומר פונ' הצפיפויות יהיו חופפות.

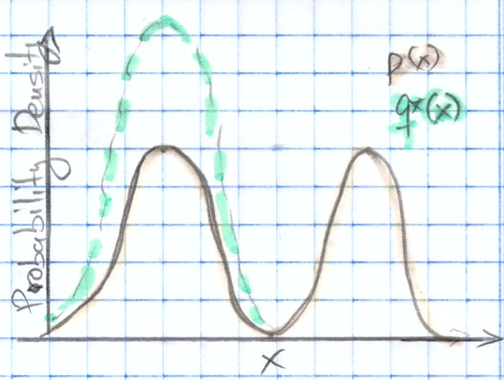
עבור $q(x)=0$ מתקבל שהפונ' מתבדרת:

$$\log\left(\frac{p(x)}{q(x)}\right) \xrightarrow[q(x)\to 0]{} \infty$$

ולכן נשאף עבור $p(x) \gneqq 0$ יתקבל $q(x)>0$.

(b) $q^* = \underset{q}{\arg\min}\, KL(q\|p)$

$= \underset{q}{\arg\min}\left\{E_{x\sim p}\left[\log\left(\frac{q^*(x)}{p(x)}\right)\right]\right\}$



במקרה (b) אנו שואפים למזער את הביטוי: $\log\left(\frac{q^*(x)}{p(x)}\right)$.
בשונה מ-(a), כאן נדרוש כי עבור $p(x)=0$ גם $q^*(x)=0$, ולכן במקרה זה $q^*(x)$ מנסה שלא להתרחק מעבר לתמך של אחד הגאוסיינים ב-$p(x)$.