Jake Schinto

Homework 6

K-Means, GMM, and EM algorithm

1. $r_{nk} = \begin{cases} 1 & \text{if } k = \arg\min_j \|x_n - \mu_j\|^2 \\ 0 & \text{otherwise} \end{cases}$ $\quad \mu_j = \{2, 4\}$

$n=1$: $x_n = 1 \to k=1$ $\quad \|1-2\|^2 = 1$ $\quad r_{nk}=1$
$\quad\quad\quad\quad\quad\quad \to k=2$ $\quad \|1-4\|^2 = 9$ $\quad r_{nk}=0$

$n=2$: $x_n = 2 \to k=1$ $\quad \|2-2\|^2 = 0$ $\quad r_{nk}=1$
$\quad\quad\quad\quad\quad\quad \to k=2$ $\quad \|2-4\|^2 = 4$ $\quad r_{nk}=0$

$n=3$: $x_n = 3 \to k=1$ $\quad \|3-2\|^2 = 1$ $\quad r_{nk}=1$ or $r_{nk}=0$
$\quad\quad\quad\quad\quad\quad \to k=2$ $\quad \|3-4\|^2 = 1$ $\quad r_{nk}=0$ $\quad\quad r_{nk}=1$

$n=4$: $x_n = 4 \to k=1$ $\quad \|4-2\|^2 = 4$ $\quad r_{nk}=0$
$\quad\quad\quad\quad\quad\quad \to k=2$ $\quad \|4-4\|^2 = 0$ $\quad r_{nk}=1$

$\mu_j = \left\{\frac{1+2+3}{3}, \frac{4}{1}\right\} = \{2, 4\}$ or $\mu_j = \left\{\frac{1+2}{2}, \frac{3+4}{2}\right\} = \left\{\frac{3}{2}, \frac{7}{2}\right\}$

$n=1$: $x_n = 1 \to k=1$ $\quad \|1-\frac{3}{2}\|^2 = 0.25$ $\quad r_{nk}=1$
$\quad\quad\quad\quad\quad\quad \to k=2$ $\quad \|1-\frac{7}{2}\|^2 = 6.25$ $\quad r_{nk}=0$

$n=2$: $x_n = 2 \to k=1$ $\quad \|2-\frac{3}{2}\|^2 = 0.25$ $\quad r_{nk}=1$
$\quad\quad\quad\quad\quad\quad \to k=2$ $\quad \|2-\frac{7}{2}\|^2 = 2.25$ $\quad r_{nk}=0$

$n=3$: $x_n = 3 \to k=1$ $\quad \|3-\frac{3}{2}\|^2 = 2.25$ $\quad r_{nk}=0$
$\quad\quad\quad\quad\quad\quad \to k=2$ $\quad \|3-\frac{7}{2}\|^2 = 0.25$ $\quad r_{nk}=1$

$n=4$: $x_n = 4 \to k=1$ $\quad \|4-\frac{3}{2}\|^2 = 6.25$ $\quad r_{nk}=0$
$\quad\quad\quad\quad\quad\quad \to k=2$ $\quad \|4-\frac{7}{2}\|^2 = 0.25$ $\quad r_{nk}=1$

no change in r_{nk}

Cluster outcomes: $\{1,2,3\}, \{4\}$ and $\{1,2\}, \{3,4\}$
μ_j: $\{2, 4\}$ and $\left\{\frac{3}{2}, \frac{7}{2}\right\}$

$$J_1 = \sum_{i=1}^{M} \sum_{j=1}^{K} r_{ij} (x^{(i)} - \mu_j)^2 = (1-2)^2 + (2-2)^2 + (3-2)^2 + (4-4)^2 = 2$$

$$J_2 = (1-\tfrac{3}{2})^2 + (2-\tfrac{3}{2})^2 + (3-\tfrac{7}{2})^2 + (4-\tfrac{7}{2})^2 = 1$$

2.

$$\mathcal{N}(X|\mu,\Sigma) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right\}$$

$$\frac{\partial}{\partial\mu} \ln p(X|\mu,\Sigma) = \sum_{n=1}^{N} \Sigma^{-1}(x_n-\mu)$$

$$\mu_{ML} = \frac{1}{N}\sum_{n=1}^{N} x_n$$

$$\ln p(X|\pi,\mu,\Sigma) = \sum_{n=1}^{N} \ln \underbrace{\left\{\sum_{j=1}^{K} \pi_j \mathcal{N}(x_n|\mu_j,\Sigma_j)\right\}}$$

$$\frac{\partial}{\partial\mu_j} \rightarrow \sum_{n=1}^{N} \frac{\partial}{\partial u}\ln(u) \cdot \frac{\partial u}{\partial\mu_j} \rightarrow u = \text{ร}$$

$$\frac{\partial u}{\partial\mu_j} = \sum_{j=1}^{K} \frac{\partial}{\partial\mu_j}\left(\pi_j \mathcal{N}(x_n|\mu_j,\Sigma_j)\right)$$

$$\cancel{\frac{\partial\pi_j}{\partial\mu_j}\cdot\mathcal{N}(x_n|\mu_j,\Sigma_j)} + \pi_j \sum_{n=1}^{N}\left(\Sigma^{-1}(x_n-\mu_j)\cdot\mathcal{N}(x_n|\mu_j,\Sigma_j)\right) = \frac{\partial u}{\partial\mu_j}$$

$$\frac{\partial}{\partial\mu_j} \Rightarrow \sum_{n=1}^{N} \frac{\pi_j \cdot \mathcal{N}(x_n|\mu_j,\Sigma_j)}{\sum_t \pi_t \mathcal{N}(x_t|\mu_t,\Sigma_t)} \cdot \Sigma^{-1}(x_n-\mu_j) \quad \leftarrow \frac{1}{u}\cdot\frac{\partial u}{\partial\mu_j}$$

3. a. Jensen's inequality:

$$f\left(\sum x\, p(x)\right) \leq \quad f(x)\,(x)$$

applied to $KL(p\|q)$:

$$-\sum q(x) \ln\left\{\frac{p(x)}{q(x)}\right\} \geq -\ln \sum p(x)$$

where $\sum p(x) = 1$

So, $-\sum q(x) \ln\left\{\frac{p(x)}{q(x)}\right\} \geq 0$ or non-negative

b. $L(q,\theta) + KL(q||p)$

$$= \sum_z q(z) \ln\left\{\frac{p(x,z|\theta)}{q(z)}\right\} - \sum_z q(z) \ln\left\{\frac{p(z|x,\theta)}{q(z)}\right\}$$

$$= \sum_z q(z) \left(\ln\left\{\frac{p(x,z|\theta)}{q(z)}\right\} - \ln\left\{\frac{p(z|x,\theta)}{q(z)}\right\} \right)$$

$$= \sum_z q(z) \left(\ln\left\{\frac{p(x,z|\theta)}{q(z)} \cdot \frac{q(z)}{p(z|x,\theta)}\right\} \right)$$

$$= \sum_z q(z) \underbrace{\left(\ln(p(x,z|\theta)) - \ln(p(z|\cancel{x},\theta)) \right)}$$

$$= \sum_z q(z) \cdot \ln(p(x|\theta)) = \ln(p(x|\theta)) \cdot \underbrace{\sum_z q(z)}_{1}$$

$$= \ln(p(x|\theta)) = L(q,\theta) + KL(q||p)$$

PCA

1. $\Sigma = \frac{1}{m} \sum_i^m x^{(i)} x^{(i)T} \rightarrow u_j^T \Sigma u_j$

$$= u_j^T \cdot \frac{1}{m} \sum_i^m x^{(i)} x^{(i)T} \cdot u_j \rightarrow u_j^T \cdot x^{(i)} = x^{(i)T} \cdot u_j$$

$$= \frac{1}{m} \sum_i^m (u_j^T \cdot x^{(i)} \cdot x^{(i)T} \cdot u_j) = \frac{1}{m} \sum_i^m (u_j^T x^{(i)})^2$$

2.a.

$$J = \frac{1}{m} \sum_i^m ||x^{(i)} - \tilde{x}^{(i)}||^2 = \frac{1}{m} \sum_i^m ||x^{(i)} - \sum_j^M z_i u - \sum_{j=M}^n b u||^2$$

$$\frac{\partial}{\partial z} \rightarrow \cancel{x^{(i)2}} - x^{(i)} \sum_j^M zu + \left(\sum_j^M zu\right)^2 + \left(\sum_j^M zu\right)\left(\sum_j^n bu\right) + \cancel{\left(\sum_j^n bu\right)^2}$$

$$= -x^{(i)} \left(\cancel{\sum_j^M u_j}\right) + 2\left(\sum_j^M zu\right)\left(\cancel{\sum_j^M u}\right) + \left(\sum_j^n bu\right)\left(\cancel{\sum_j^M u}\right) = 0$$

$j \leq M$

$$2\left(\sum_j^M zu\right) = -\cancel{\left(\sum_j^n bu\right)} + x^{(i)}$$

$$z_{ij} = x_i^T u_j$$

b. $\frac{\partial}{\partial b} \rightarrow -\bar{x}\left(\sum_{j}^{m} bu\right) + \left(\sum_{j}^{k} bu\right)^2 + \left(\sum_{j}^{k} bu\right)\left(\sum_{j}^{k} zu\right)$

$\rightarrow -\bar{x}\left(\cancel{\sum_{j}^{k} u}\right) + 2\left(\sum_{j}^{k} bu\right)\left(\cancel{\sum_{j}^{k} u}\right) + \left(\sum_{j}^{k} zu\right)\left(\cancel{\sum_{j}^{k} u}\right) = 0$

$b_j = \bar{x}^T u_j$

c.

$u_j^T \Sigma u_j = \frac{1}{m} \sum_{i=1}^{m} \left(u_j^T x^{(i)}\right)^2$

$\sum_{j=m}^{k} \left(\frac{1}{m} \sum_{i=1}^{m} \left(u_j^T x^{(i)}\right)^2\right) = \frac{1}{m} \sum_{i=1}^{m} \left(\sum_{j=m}^{k} \left(u_j^T x^{(i)}\right)^2\right)$

$= \frac{1}{m} \sum_{i=1}^{m} \left(\sum_{j=m}^{k} \left(x_n^T u_i - \bar{x}^T u_i\right)^2\right) = \frac{1}{m} \sum_{i=1}^{m} \left(x - \tilde{x}\right)^2 = J$

Neural Networks

1. $\hat{y} = \frac{\exp(z_j)}{\sum_t \exp(z_t)}$ where $t = \{0, \dots, K-1\}$

therefore $\hat{y}$ must be over the same set as t.

$\ell(y, \hat{y}) = -\sum_{j=0}^{K-1} y_j \ln \hat{y}$ where $\hat{y} = \frac{\exp(z_j)}{\sum_t \exp(z_t)}$

Sum of all js $\rightarrow \sum_j \hat{y} = 1$

$= -\sum_{j=0}^{K-1} y_j \left(\ln(\exp(z_j)) - \ln\left(\sum_t \exp(z_t)\right)\right)$

$\hat{y}$ is a probability distribution

where exp is non-negative and $\sum_t \exp(z_t) \geq \exp(z_j)$ because z_j is contained in z_t summation.

- Therefore $\ln \hat{y}$ will be negative from the difference and y_j is positive, making $y_j \ln \hat{y} \leq 0$
  - This makes the $\ell(y, \hat{y})$ non-negative because of the leading negative multiplied by a negative term

$\frac{\partial}{\partial z_j} \rightarrow -y_j \sum_{j=0}^{K-1} \left(1 - \underbrace{\frac{1}{\sum_t \exp(z_t)} \exp(z_j)}_{\hat{y}_j}\right) = \hat{y}_j - y_j$

2. $a^{[1]} = z^{[1]} = w^{[1]}x + b^{[1]}$

$a^{[2]} = z^{[2]} = w^{[2]}(w^{[1]}x + b^{[1]}) + b^{[2]}$

$= (w^{[2]}w^{[1]})x + (w^{[2]}b^{[1]}) + b^{[2]}$

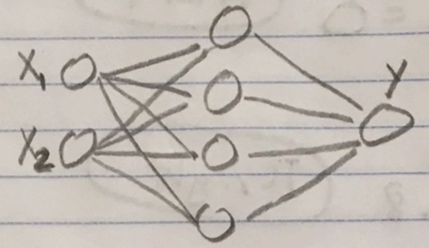
set $b^{[1]} = b^{[2]} = 0$

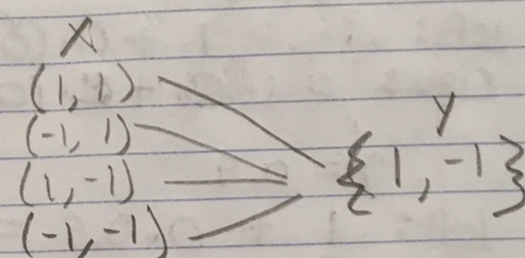
$w^{[2]} = u^T$ and $w^{[1]} = u$

it now takes the form $\hat{y} = \sum_{j=1}^{m} u_j^T x u_j$

$\ell = \sum_{i=1}^{m} \| x^{(i)} - \sum_{j=1}^{m} u^T x u \|^2$

3.





all 4 possibilities are covered since they are all boolean options

Reinforcement Learning

1. $\frac{1}{4}(R + \gamma V^\pi(s' = n)) + \frac{1}{4}(R + \gamma V^\pi(s' = e))$

$+ \frac{1}{4}(R + \gamma V^\pi(s' = s)) + \frac{1}{4}(R + \gamma V^\pi(s' = w))$

where n, e, s, w are the 4 directions

$R = 0$ for each step

$\frac{1}{4} \cdot \gamma \cdot (2.3 + 0.4 - 0.4 + 0.7) = \frac{0.9}{4}(3) = 0.675$

≈ 0.7

2. it would be the average of the actions

$V^\pi(s) = E\{R \mid s_{state} = s\}$ $\quad P\{a_t = a \mid s_t = s\}$

$= \sum_a E\{R \mid s_t = s, a_t = a\}\, \pi(s,a)$

$V^\pi(s) = \sum_a Q^\pi(s,a)\,\pi(s,a)$

3. $\gamma = 0$

left: $1 + 0\cdot(\overset{\gamma}{8} - 0) = 1$ $\quad \pi_{left}$

right: $0 + 2\cdot(0 - 0) = 0$

$\gamma = 0.9$

left: $1 + 0 \cdot 0.9 = 1$

right: $0 + 2 \cdot 0.9 = 1.8$ $\quad \pi_{right}$

$\gamma = 0.5$

left: $1 + 0 \cdot 0.5 = 1$

$0 + 2 \cdot 0.5 = 1$ $\quad$ both optimal