Vele Tosevski

1002174657

CSC 411

March 20, 2019

HW5 Write-Up

1. a)

   Average Conditional Likelihood (Train Data): 28.686580

```
[[ 33.92402221 -88.45111834 -16.8932622 -25.83617441 -95.99459913 -28.0469699 -33.43930375 -128.00676019 -40.12341711 -120.7505451 ]
 [ -68.22979333 43.8330528 -8.62313018 -41.79892253 -6.16921495 -42.54976231 -22.89991813 -33.06720503 4.20662396 -19.78080586]
 [ -48.25937897 -55.92309386 16.88110064 -28.9959108 -65.50319452 -47.8188899 -54.71585182 -84.2803214 -35.97571408 -79.85123569]
 [ -37.76535226 -60.50596233 -11.52881147 25.34749836 -80.23781057 -5.942891 -61.89235394 -74.20703712 -7.77750941 -63.81771045]
 [ -63.57874438 -38.1128814 -27.37434988 -43.54955558 24.20650818 -37.62118033 -38.11452022 -49.30644691 -19.02665868 -3.93941094]
 [ -31.92250461 -45.89294551 -23.0214228 0.2051562 -56.8635482 24.12660812 -36.96891002 -84.4180825 -2.6016538 -53.5127435 ]
 [ -16.08499354 -65.52399532 -17.22797111 -38.51566436 -59.26386047 -19.90746009 29.24659472 -165.40709904 -29.97466814 -126.96745943]
 [ -47.34810481 -22.74497808 0.17304175 -13.15842334 -22.22716495 -39.09215712 -80.81237908 38.53051509 -7.68591751 9.77505843]
 [ -51.89004965 -33.31890348 -15.94135661 -21.54837236 -45.81703868 -18.8929108 -47.39728619 -66.1336154 20.02956548 -37.4736449 ]
 [ -65.1618034 -46.64166345 -24.55400204 -32.3243134 -3.74088864 -39.22988767 -85.57166051 -7.67267147 -5.99624188 30.74033154]]
```

   Average Conditional Likelihood (Test Data): 27.486200

```
[[ 33.21271943 -85.10474932 -14.92668303 -25.03919668 -92.24055178 -28.41301758 -31.98166934 -121.11671514 -38.9435702 -115.68602049]
 [ -65.98273597 43.75667838 -7.89988668 -40.78305048 -4.21465107 -41.85439817 -19.49559598 -27.26307905 5.51581709 -17.43312634]
 [ -48.40011136 -56.79648782 15.76987273 -32.02961238 -64.05056948 -48.58909001 -52.90835733 -87.94971095 -36.4708241 -81.13238684]
 [ -36.61811894 -59.0998393 -10.81252291 23.18843659 -79.32152079 -5.49772638 -60.31175001 -75.53668363 -8.21771127 -64.68566461]
 [ -63.6224916 -38.31251625 -27.68532145 -42.79967361 23.16501442 -36.98680559 -38.8213785 -48.45629329 -19.54241646 -2.7608024 ]
 [ -28.61602964 -40.69825848 -20.36105698 3.672266 -55.92112179 24.03785909 -32.49333123 -81.85722903 -1.63104965 -53.09672336]
 [ -17.94182886 -70.13173546 -19.32409691 -40.17621616 -61.330969 -20.45038953 26.35234161 -169.74575806 -30.4058043 -128.24679529]
 [ -48.50416378 -22.52592109 -0.48298009 -13.97483349 -20.82250406 -38.86921001 -81.1159767 37.69449875 -8.12088509 9.78420061]
 [ -53.05155379 -34.54053281 -17.60860408 -22.40460407 -46.75337517 -19.61290705 -49.00099401 -68.0970773 18.6596498 -38.65401494]
 [ -64.02644638 -46.67289562 -22.92041583 -30.60657673 -4.95359756 -38.41795983 -86.48303501 -6.37184554 -5.88093645 29.02492862]]
```

   b) Accuracies -> Train: 98.143% (Rounded)

   Test: 97.275%

   c) The following figure is the figure of eigenvectors of the Covariance matrices for each digit:

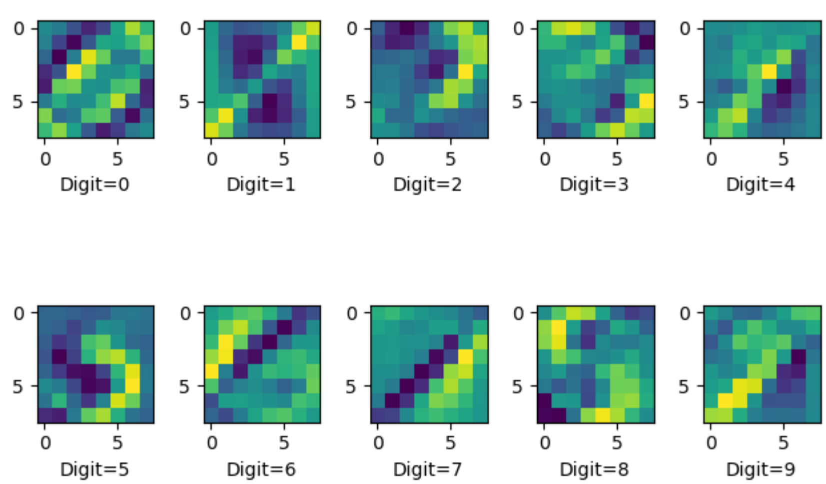

Figure 1: Covariance Eigenvalues for each Digit (Axes are pixel number)

2. a). Need to find $P(\theta | D)$.

$$p(\theta|D) = \frac{p(D|\theta)\, p(\theta)}{p(D)} \propto p(D|\theta)\, p(\theta) \qquad ①$$

$$p(D|\theta) = p(x_1, x_2, \ldots, x_N | \theta)$$

$$= \frac{p(x_1, x_2, \ldots, x_N, \theta)}{p(\theta)}$$

Assuming conditional independence

$$= \frac{\cancel{p(\theta)}\, p(x_1|\theta)\, p(x_2|\theta) \cdots p(x_N|\theta)}{\cancel{p(\theta)}}$$

$$= p(x_1|\theta)\, p(x_2|\theta) \cdots p(x_N|\theta)$$

$$= \prod_{i=1}^{N} p(x_i|\theta) \qquad ②$$

$$p(x|\theta) = \theta_1^{x_1}\theta_2^{x_2} \cdots \theta_K^{x_K} = \prod_{j=1}^{K} \theta_j^{x_j} \qquad ③$$

③ → ②

$$p(D|\theta) = \prod_{i=1}^{N}\left(\prod_{j=1}^{K} \theta_j^{x_j^{(i)}}\right) \qquad ④$$

$$p(\theta) \propto \theta_1^{a_1-1}\theta_2^{a_2-1} \cdots \theta_K^{a_K-1} = \prod_{i=1}^{K} \theta_i^{a_i-1} \qquad ⑤$$

④, ⑤ → ①

$$p(\theta|D) \propto p(D|\theta)\, p(\theta) = \boxed{\left[\prod_{i=1}^{N}\left(\prod_{j=1}^{K} \theta_j^{x_j^{(i)}}\right)\right]\left[\prod_{i=1}^{K} \theta_i^{a_i-1}\right]}$$

$$p(\theta \mid D) \propto \left[\prod_{n=1}^{N}\left(\prod_{k=1}^{K} \theta_k^{x_k^{(n)}}\right)\right]\left[\prod_{k=1}^{K} \theta_k^{a_k-1}\right]$$

$$= \left[\prod_{k=1}^{K}\left(\prod_{n=1}^{N} \theta_k^{x_k^{(n)}}\right)\right]\left[\prod_{k=1}^{K} \theta_k^{a_k-1}\right]$$

$$= \left[\prod_{k=1}^{K}\left(\theta_k^{x_k^{(1)}+x_k^{(2)}+\dots+x_k^{(N)}}\right)\right]\left[\prod_{k=1}^{K} \theta_k^{a_k-1}\right]$$

$$= \prod_{k=1}^{K}\left(\theta_k^{N_k}\,\theta_k^{a_k-1}\right) \quad \text{since } \sum_{n=1}^{N} x_k^{(n)} = N_k$$

$$= \# \text{ of times } k = 1$$

$$= \prod_{k=1}^{K} \theta^{(N_k+a_k-1)} \sim \text{Dirichlet}(N_1+a_1, \dots, N_K+a_K) \quad (1)$$

We also know that $p(x \mid \theta) = p(D \mid \theta) = \prod_{k=1}^{K} \theta_k^{x_k}$

So, $p(D' \mid \theta) = \prod_{k=1}^{K} \theta_k^{x'_k} = \theta_k$ where $x'_k = k$ (2)

$$p(D' \mid D) = \sum_{k=1}^{K} \overset{(1)}{p(\theta_k \mid D)}\, \overset{(2)}{p(D' \mid \theta_k)}$$

$$= \sum_{k=1}^{K} \theta_k \prod_{i=1}^{K} \theta^{(N_i+a_i-1)}$$

$$= \sum_{k=1}^{K} \theta_k \, \text{Dirichlet}(N_1+a_1, \dots, N_K+a_K)$$

$$= E[\theta_k] \text{ over Dirichlet}(N_1+a_1, \dots, N_K+a_K)$$

$$= \boxed{\frac{a_k+N_k}{\sum_{k'}(a_{k'}+N_{k'})} \quad \text{where } N_{k'} = \sum_{n=1}^{N} x_{k'}^{(n)}}$$

2 . b).

$$\hat{\theta}_{MAP} = \underset{\theta}{\arg\max} \left(\log p(\theta) + \log p(D|\theta) \right)$$

Recall: $p(\theta) = \prod_{k=1}^{K} \theta_k^{a_k - 1}$

$$p(D|\theta) = \prod_{k=1}^{K} \prod_{n=1}^{N} \theta_k^{x_k^{(n)}} = \prod_{k=1}^{K} \theta_k^{N_k} \quad \text{where } N_k = \sum_{n=1}^{N} x_k^{(n)}$$

$$\log(p(\theta)) + \log(p(D|\theta))$$

$$= \log \prod_{k=1}^{K} \theta_k^{a_k - 1} + \log \prod_{k=1}^{K} \theta_k^{N_k}$$

$$= \sum_{k=1}^{K} \left(\log(\theta_k^{a_k-1}) + \log(\theta_k^{N_k}) \right)$$

$$= \sum_{k=1}^{K} (a_k - 1)\log\theta_k + N_k \log\theta_k$$

$$= \sum_{k=1}^{K} (a_k - 1 + N_k)\log(\theta_k)$$

$$\frac{\partial}{\partial\theta_k}(\hat{\theta}_{MAP}) = \sum_{k=1}^{K} \frac{\partial}{\partial\theta_k}(a_k - 1 + N_k)\log(\theta_k)$$

$$= \frac{a_k + N_k - 1}{\theta_k} = 0$$

$\therefore$ Let's try Lagrange to solve for θ_k.

$$\nabla f(\theta_k) = \frac{a_k + N_k - 1}{\theta_k}$$

$$g(\theta) = 1 - \sum_{k=1}^{K} \theta_k$$

$$\nabla g(\theta_k) = -1$$

$\rightarrow$

$$\nabla f = \lambda \nabla g$$

$$\frac{a_k + N_k - 1}{\theta_k} = -\lambda \rightarrow \theta_k = \frac{1 - a_k - N_k}{\lambda}$$

$$\therefore\ 1 - \sum_{k=1}^{K} \theta_k = 0$$

$$1 - \sum_{k=1}^{K} \frac{1 - a_k - N_k}{\lambda} = 0$$

$$\lambda - \sum_{k=1}^{K}(1-a_k-N_k)=0$$

$$\lambda = \sum_{k=1}^{K}(1-a_k-N_k)$$

$$\lambda = K - N - \sum_{k=1}^{K} a_k \quad \text{since } \sum_{k=1}^{K} N_k = N$$

$$\therefore \quad \theta_k = \frac{1-a_k-N_k}{\lambda} = \frac{1-a_k-N_k}{K-N-\sum_{k'=1}^{K} a_{k'}}$$

$$\therefore \quad \boxed{\hat{\theta}_{MAP} = \left(\frac{1-a_1-N_1}{K-N-\sum_{k=1}^{K} a_k}, \cdots, \frac{1-a_k-N_k}{K-N-\sum_{k'=1}^{K} a_{k'}}, \cdots, \frac{1-a_K-N_K}{K-N-\sum_{k=1}^{K} a_k} \right)}$$

3.a). We know,

$$z \sim \mathcal{N}(0, 1)$$
$$x|z \sim \mathcal{N}(zu, \Sigma)$$

for z scalar

From the appendix,

$$p(z) = \mathcal{N}(z \mid \mu, \Lambda^{-1})$$
$$p(x|z) = \mathcal{N}(x \mid Az+b, L^{-1})$$

Taking the univariate version, (since it matches for 1 dimension)

$$\left.\begin{array}{l} p(z) = \mathcal{N}(z \mid \mu, \Lambda^{-1}) \\ \mathcal{N}(0,1) = \mathcal{N}(z \mid \mu, \Lambda^{-1}) \end{array}\right\} \rightarrow \mu = 0, \ \Lambda^{-1} = 1 = \Lambda \qquad (1)$$

$$\left.\begin{array}{l} p(x|z) = \mathcal{N}(x \mid Az+b, L^{-1}) \\ \mathcal{N}(zu, \Sigma) = \mathcal{N}(x \mid Az+b, L^{-1}) \end{array}\right\} \rightarrow \begin{array}{l} Az+b = zu \rightarrow A = u, \ b = 0 \\ L^{-1} = \Sigma \\ L = \Sigma^{-1} \end{array} \qquad (2)$$

Now we can solve for $p(z|x;\theta)$.

Converting ...

$$p(z|x;\theta) = \mathcal{N}(z \mid C(A^T L(x-b) + \Lambda\mu), C)$$

$$C = (\Lambda + A^T L A)^{-1}$$
$$= (1 + u^T \Sigma^{-1} u)^{-1}$$

(1), (2)

So,

$$p(z|x;\theta) = \mathcal{N}(z \mid (1+u^T\Sigma^{-1}u)^{-1}(u^T\Sigma^{-1}x + 0), (1+u^T\Sigma^{-1}u)^{-1})$$

$$= \boxed{\mathcal{N}(z \mid (1+u^T\Sigma^{-1}u)^{-1}u^T\Sigma^{-1}x, (1+u^T\Sigma^{-1}u)^{-1})}$$

Now,

$$m = E[z \mid x; \theta]$$

$$= E\left[\mathcal{N}\left(z \mid (1+u^T\Sigma^{-1}u)^{-1}u^T\Sigma^{-1}x,\ (1+u^T\Sigma^{-1}u)^{-1}\right)\right]$$

$$= E\left[\mathcal{N}(z \mid \mu, \sigma^2)\right]$$

Since it is the expected value of a normal distribution, then $E(N)$ is simply the mean of the normal distribution.

$$= \boxed{(1+u^T\Sigma^{-1}u)^{-1}u^T\Sigma^{-1}x}$$

check: $(1+\underbrace{u^T}_{1\times d}\underbrace{\Sigma^{-1}}_{d\times d}\underbrace{u}_{d\times 1})^{-1}$ → 1×1 ✓ univariate

$(\underbrace{u^T}_{1\times d}\underbrace{\Sigma^{-1}}_{d\times d}\underbrace{x}_{d\times 1})$ → 1×1 ✓ univariate

Univariate ∴ ✓

Lin. func. of x ✓

$$s = E[z^2 \mid x; \theta]$$

$$= Var(z \mid x; \theta) + E^2[z \mid x; \theta]$$

$$= (1+u^T\Sigma^{-1}u)^{-1} + \left[(1+u^T\Sigma^{-1}u)^{-1}u^T\Sigma^{-1}x\right]^2$$

$$= \boxed{\underbrace{(1+u^T\Sigma^{-1}u)^{-1}}_{\text{Univariate}}\left[1 + \underbrace{(1+u^T\Sigma^{-1}u)^{-1}}_{\text{Univariate}}\underbrace{(u^T\Sigma^{-1}x)^2}_{\text{Univariate}}\right]}$$

Univariate ✓

3. b). $\mathbb{E}_q[p(z^{(n)}, x^{(n)}; \theta)]$ $\mathbb{E}_q[p(x^{(n)} | z^{(n)}; \theta)\, p(z^{(n)}; \theta)]$ (0)

$$\log(p(z^{(n)}, x^{(n)}; \theta)) = \underbrace{\log p(x^{(n)} | z^{(n)}; \theta)}_{\sim N(zu, \Sigma)} + \underbrace{\log p(z^{(n)}; \theta)}_{\sim N(0,1)}$$

$\log(p(x^{(n)} | z^{(n)}; \theta))$

$$= \log\left[(2\pi)^{-D/2} |\Sigma|^{-1/2} \exp\left(\frac{-1}{2}(x^{(n)} - z^{(n)}u)^T \Sigma^{-1} (x^{(n)} - z^{(n)}u)\right)\right]$$

$$= \frac{-D}{2}\log(2\pi) - \frac{1}{2}\log|\Sigma| - \frac{1}{2}\left[x^{(n)T}\Sigma^{-1}x^{(n)} - x^{(n)T}\Sigma^{-1}z^{(n)}u - z^{(n)T}u^T\Sigma^{-1}x^{(n)} + z^{(n)T}u^T\Sigma^{-1}z^{(n)}u\right]$$

$$= \left[\frac{-D}{2}\log(2\pi) - \frac{1}{2}\log|\Sigma| - \frac{1}{2}x^{(n)T}\Sigma^{-1}x^{(n)} + \frac{1}{2}z^{(n)}x^{(n)T}\Sigma^{-1}u + \frac{1}{2}z^{(n)}u^T\Sigma^{-1}x^{(n)} - \frac{1}{2}(z^{(n)})^2 u^T\Sigma^{-1}u\right]$$

$\log p(z^{(n)}; \theta)$ (1)

$$= \log\left[(2\pi)^{-1/2}\, 1 \cdot \exp\left(\frac{-1}{2}(z^{(n)} - 0)^T \cdot 1 \cdot (z^{(n)} - 0)\right)\right]$$

$$= \frac{-1}{2}\log(2\pi) - \frac{1}{2}z^{(n)T}z^{(n)}$$

$$= \frac{-1}{2}\log(2\pi) - \frac{1}{2}(z^{(n)})^2 \quad (2)$$

$$\log p(z^{(n)}, x^{(n)}; \theta) = \log p(x^{(n)} | z^{(n)}; \theta) + \log p(z^{(n)}, \theta)$$

(1) (2)

$$= \left(\frac{D+1}{2}\right)(-\log(2\pi)) - \frac{1}{2}\log|\Sigma| - \frac{1}{2}x^{(n)T}\Sigma^{-1}x^{(n)} + \frac{1}{2}z^{(n)}x^{(n)T}\Sigma^{-1}u + \frac{1}{2}z^{(n)}u\Sigma^{-1}x^{(n)}$$

$$= -\frac{(D+1)}{2}\log(2\pi) - \frac{1}{2}\log|\Sigma| - \frac{1}{2}x^{(n)T}\Sigma^{-1}x^{(n)} + \frac{1}{2}x^{(n)T}\Sigma^{-1}u\, m^{(n)} \quad -\frac{1}{2}(z^{(n)})^2(1 + u^T\Sigma^{-1}u)$$

$$+ \frac{1}{2}u\Sigma^{-1}x^{(n)}m^{(n)} - \frac{1}{2}s^{(n)}(1 + u^T\Sigma^{-1}u)$$

$$\frac{\partial}{\partial u} = \frac{1}{N}\sum_{n=1}^{N} (0) \cdot \left[\frac{1}{2}x^{(n)T}\Sigma^{-1}m^{(n)} + \frac{1}{2}\Sigma^{-1}x^{(n)}m^{(n)} - \frac{1}{2}s^{(n)}u^T\Sigma^{-1}\right]$$

$$= \frac{1}{N}\sum_{n=1}^{N} \cancel{(0)} \cdot \left[x^{(n)T}\Sigma^{-1}m^{(n)} - s^{(n)}u^T\Sigma^{-1}\right] = 0$$

$$\sum_{n=1}^{N} x^{(n)T}\cancel{\Sigma^{-1}}m^{(n)} = \sum_{n=1}^{N} s^{(n)}u^T\cancel{\Sigma^{-1}}$$

$$\sum_{n=1}^{N} x^{(n)} m^{(n)} = \sum_{n=1}^{N} s^{(n)} u$$

$$\boxed{u = \frac{\sum_{n=1}^{N} x^{(n)} m^{(n)}}{\sum_{n=1}^{N} s^{(n)}}}$$