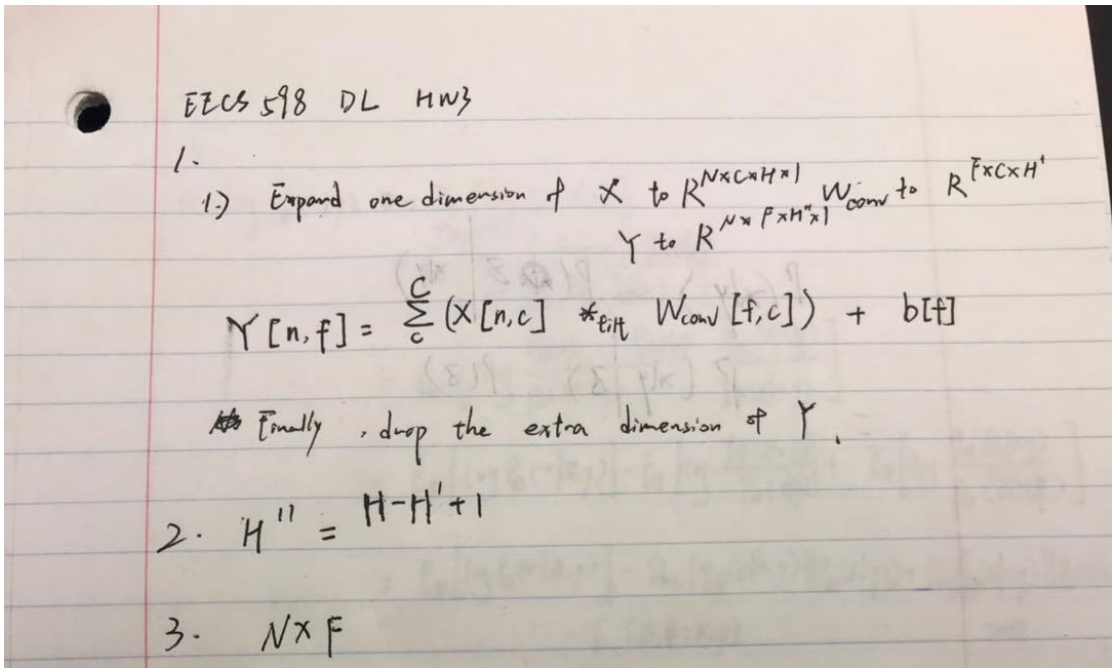
1. Text Classification with CNN

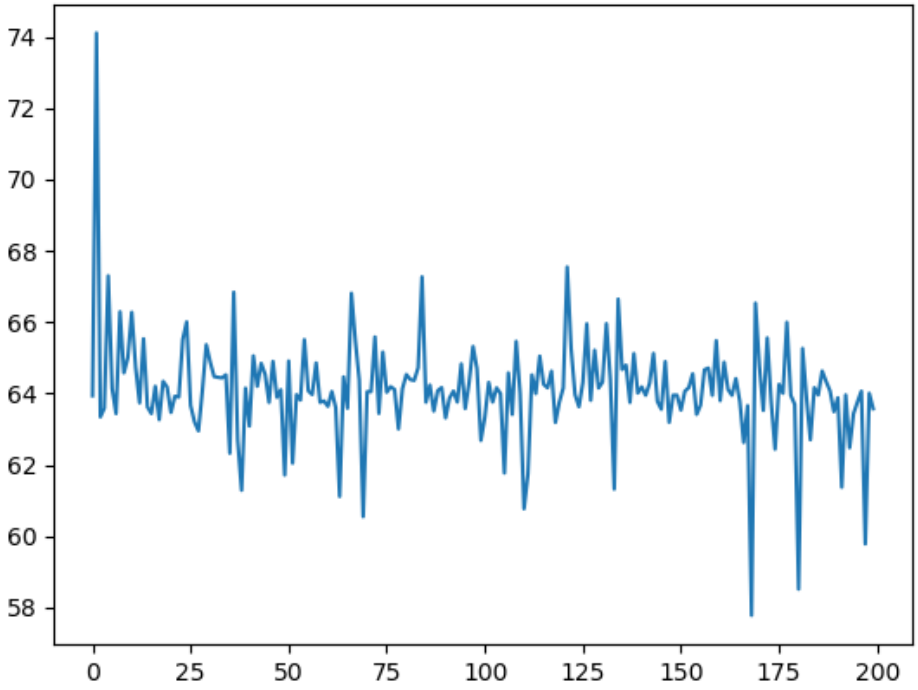


Epoch num: 50
Batch size: 500
Word Embedding: dimension 300 from scratch
Filter num: 128
Optimizer: Adam
Learning rate: 0.005
Weight decay: 0.0005
Loss function: BCE Loss

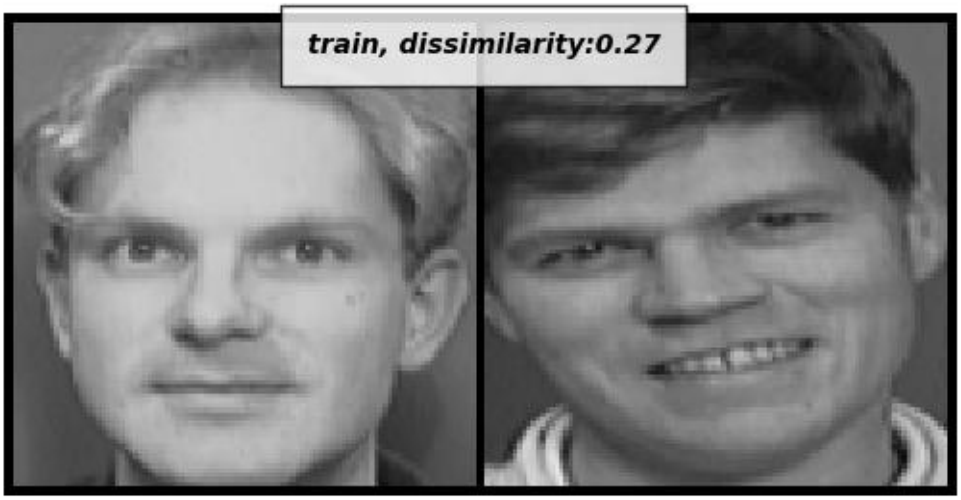
Test Accuracy	Max Pooling	Average Pooling
Kernel size 5	0.9501	0.9413
Kernel size 7	0.9509	0.9428

2. Siamese Networks for Learning Embeddings

Results are saved in code/att_result and code/lfw_result


74
72
70
68
66
64
62
60
58
0
25
50
75
100
125
150
175
200


train, dissimilarity:0.15


train, dissimilarity:0.27


train, dissimilarity:0.13


train, dissimilarity:0.10


train, dissimilarity:0.13


train, dissimilarity:0.16


train, dissimilarity:0.24


train, dissimilarity:0.10


train, dissimilarity:0.06


train, dissimilarity:0.18

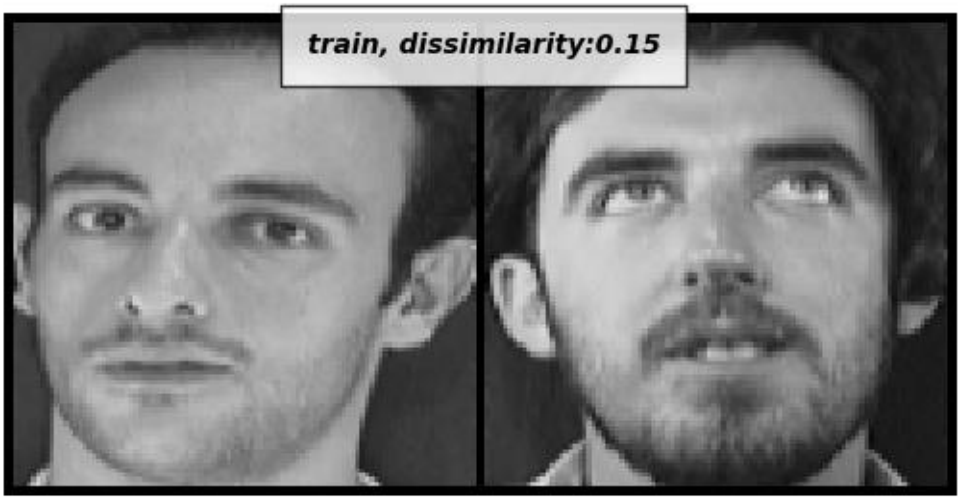

train, dissimilarity:0.12


train, dissimilarity:0.12


train, dissimilarity:0.08


train, dissimilarity:0.09


train, dissimilarity:0.21


train, dissimilarity:0.15


train, dissimilarity:0.15


train, dissimilarity:0.25


train, dissimilarity:0.10


train, dissimilarity:0.11


test, dissimilarity:0.11


test, dissimilarity:0.05


test, dissimilarity:0.00


test, dissimilarity:0.09


test, dissimilarity:0.12


test, dissimilarity:0.05


test, dissimilarity:0.10


test, dissimilarity:0.10


test, dissimilarity:0.04


test, dissimilarity:0.12


test, dissimilarity:0.10


test, dissimilarity:0.13


test, dissimilarity:0.10


test, dissimilarity:0.10


test, dissimilarity:0.10


test, dissimilarity:0.08

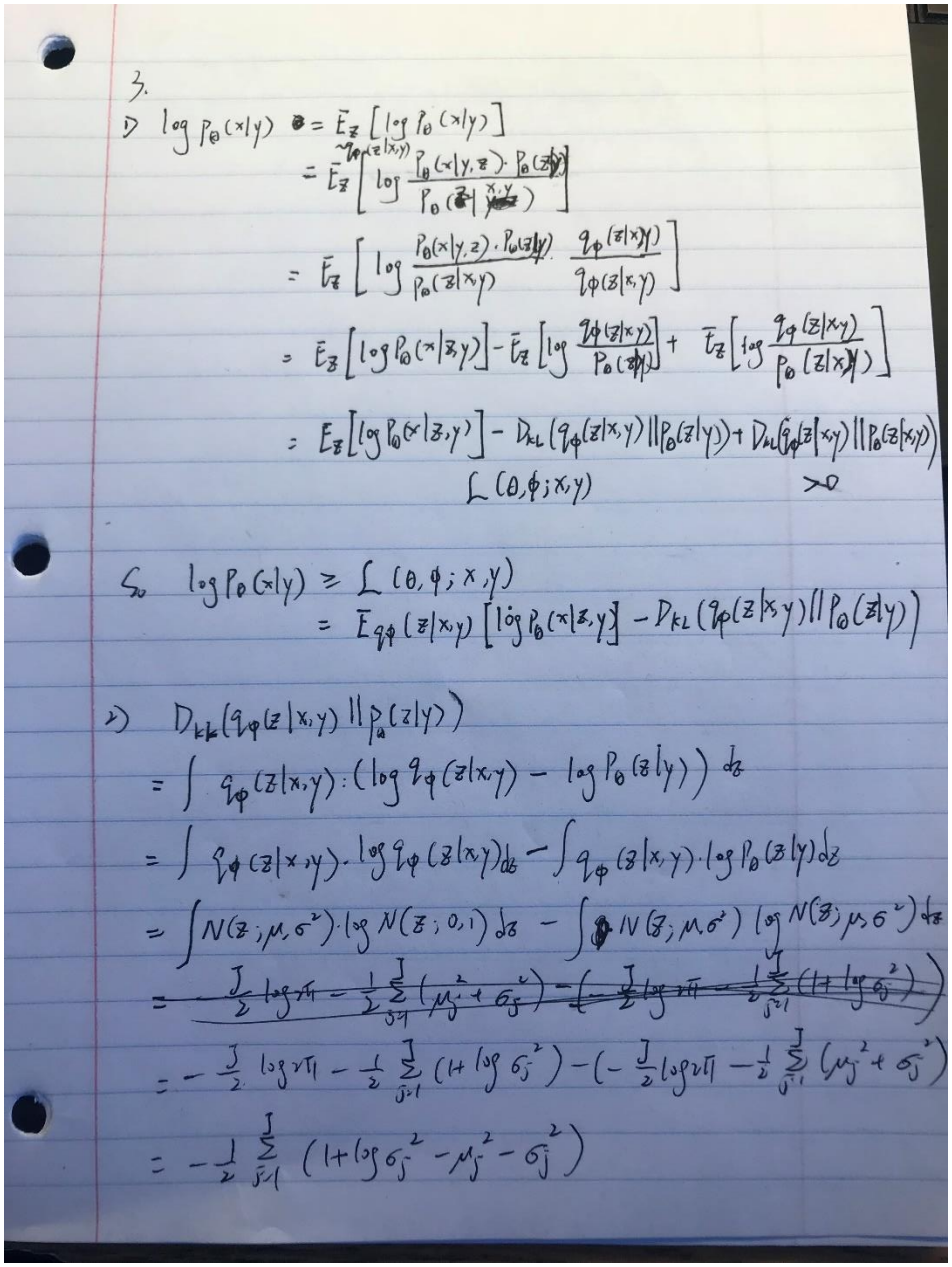

test, dissimilarity:0.10


test, dissimilarity:0.08


test, dissimilarity:0.12


test, dissimilarity:0.12

3. Conditional Variational Autoencoder



3.

1) $\log P_\theta(x|y) = E_z[\log P_\theta(x|y)]$

$= E_{z\sim q_\phi(z|x,y)}\left[\log \frac{P_\theta(x|y,z)\cdot P_\theta(z|y)}{P_\theta(z|x,y)}\right]$

$= E_z\left[\log \frac{P_\theta(x|y,z)\cdot P_\theta(z|y)}{P_\theta(z|x,y)}\cdot\frac{q_\phi(z|x,y)}{q_\phi(z|x,y)}\right]$

$= E_z[\log P_\theta(x|z,y)] - E_z\left[\log \frac{q_\phi(z|x,y)}{P_\theta(z|y)}\right] + E_z\left[\log \frac{q_\phi(z|x,y)}{P_\theta(z|x,y)}\right]$

$= E_z[\log P_\theta(x|z,y)] - D_{KL}(q_\phi(z|x,y)\|P_\theta(z|y)) + D_{KL}(q_\phi(z|x,y)\|P_\theta(z|x,y))$

$\mathcal{L}(\theta,\phi;x,y)$ $\quad >0$

So $\log P_\theta(x|y) \geq \mathcal{L}(\theta,\phi;x,y)$

$= E_{q_\phi(z|x,y)}[\log P_\theta(x|z,y)] - D_{KL}(q_\phi(z|x,y)\|P_\theta(z|y))$

2) $D_{KL}(q_\phi(z|x,y)\|P_\theta(z|y))$

$= \int q_\phi(z|x,y)\cdot(\log q_\phi(z|x,y) - \log P_\theta(z|y))\,dz$

$= \int q_\phi(z|x,y)\cdot\log q_\phi(z|x,y)\,dz - \int q_\phi(z|x,y)\cdot\log P_\theta(z|y)\,dz$

$= \int N(z;\mu,\sigma^2)\cdot\log N(z;0,1)\,dz - \int N(z;\mu,\sigma^2)\log N(z;\mu,\sigma^2)\,dz$

~~$= -\frac{J}{2}\log 2\pi - \frac{1}{2}\sum_{j=1}^{J}(\mu_j^2+\sigma_j^2) - \left(-\frac{J}{2}\log 2\pi - \frac{1}{2}\sum_{j=1}^{J}(1+\log\sigma_j^2)\right)$~~

$= -\frac{J}{2}\log 2\pi - \frac{1}{2}\sum_{j=1}^{J}(1+\log\sigma_j^2) - \left(-\frac{J}{2}\log 2\pi - \frac{1}{2}\sum_{j=1}^{J}(\mu_j^2+\sigma_j^2)\right)$

$= -\frac{1}{2}\sum_{j=1}^{J}(1+\log\sigma_j^2-\mu_j^2-\sigma_j^2)$