Gyujin Park

CSE 6740 HOMEWORK #4

1. Questions

Question 1. (a)
1. False
Let gram matrix of k_1 be K_1 and gram matrix of k_2 be K_2.
Then K_1 and K_2 are positive definite.
But K_1, K_2 positive definit does not imply that $K_1 - K_2$ positive definite.
$\therefore$ False

2.
True (textbook 296page)

3.
True

$$
\begin{aligned}
&k(u,v) = \|u-v\|^2 : \ kernel \\
&\rightarrow \gamma\|u-v\|^2 : \ kernel \\
&\rightarrow e^{\gamma\|u-v\|^2} : \ kernel
\end{aligned}
$$

(b)
If K is positive definite, then $det(K) \geq 0$.

$$\therefore k(u,u)k(v,v) \geq k(u,v)^2$$

(c)

$$
\begin{aligned}
k(u,v) &= e^{-\frac{u^Tu}{2\sigma^2}} e^{\frac{u^Tv}{\sigma^2}} e^{-\frac{v^Tv}{2\sigma^2}} \\
&= e^{-\frac{u^Tu}{2\sigma^2}} \left[\sum_{k=0}^{\infty} \frac{1}{k!} \left(\frac{u^Tv}{\sigma^2} \right)^k \right] e^{-\frac{v^Tv}{2\sigma^2}} \\
&= \sum_{k=0}^{\infty} \left[e^{-\frac{u^Tu}{2\sigma^2}} \frac{1}{k!} \left(\frac{u^Tv}{\sigma^2} \right)^k e^{-\frac{v^Tv}{2\sigma^2}} \right]
\end{aligned}
$$

$\therefore$ the Gaussian kernel can be expressed as the inner product of an infinite-dimensional feature vector.

	unnormalized	normalized
a0, b0, c0, d0	562500	0.003760969
a0, b0, c0, d1	1875000	0.012536565
a0, b0, c1, d0	2250000	0.015043878
a0, b0, c1, d1	5000000	0.03343084
a0, b1, c0, d0	1875000	0.012536565
a0, b1, c0, d1	6250000	0.04178855
a0, b1, c1, d0	6750000	0.045131634
a0, b1, c1, d1	15000000	0.10029252
a1, b0, c0, d0	2250000	0.015043878
a1, b0, c0, d1	6750000	0.045131634
a1, b0, c1, d0	9000000	0.060175512
a1, b0, c1, d1	18000000	0.120351024
a1, b1, c0, d0	5000000	0.03343084
a1, b1, c0, d1	15000000	0.10029252
a1, b1, c1, d0	18000000	0.120351024
a1, b1, c1, d1	36000000	0.240702048

FIGURE 1

Question 2. (a)

Above figure

(b)
1.
A clique is a subset of vertices of a graph such that every two distinct vertices in the clique are adjacent.
A maximal clique is a clique that cannot be extended by including one more adjacent vertex

2.
A Markov network can represent cyclic dependencies that a Bayes network cannot.

3.
Because there does not exist a parent in Markov random field.

4.

$$P(X_1, \cdots, X_n) = \frac{1}{Z} exp \left(\sum_{(i,j)\in E} \theta_{ij} X_i X_j + \sum_{i \in V} \theta_i X_i + \sum_{i \in V} \alpha_i X_i^2 \right)$$

5.
Image segmentation

Question 3. $\alpha_1(1) = \pi(1)p(H|1) = \frac{1}{6}$
$\alpha_1(2) = \pi(2)p(H|2) = \frac{1}{4}$
$\alpha_1(3) = \pi(3)p(H|3) = \frac{1}{12}$
By using forward algorithm,
$\alpha_2(1) = 0.15$
$\alpha_2(2) = \frac{17}{960}$
$\alpha_2(3) = \frac{31}{320}$

$\alpha_3(1) = \frac{597}{6400}$
$\alpha_3(2) = \frac{203}{5120}$
$\alpha_3(3) = \frac{161}{25600}$

$$\therefore P = \alpha_3(1) + \alpha_3(2) + \alpha_3(3) = 0.1392$$

Question 4. (a)
network with no hidden layer:

$$\hat{y} = \sigma(w^T x + b) = \frac{1}{1 + e^{-w^T x + b}}$$

logistic regression:

$$p(y = 1) = \frac{1}{1 + e^{-w^T x}}$$

$\therefore$ if the network has no hidden layer, that is equivalent to logistic regression.

(b)

$$l = \sum_i (y^i - \sigma(w^T z^i))^2$$

$$\sigma'(x) = \frac{e^{-x}}{(1 + e^{-x})^2} = \sigma(1 - \sigma)$$

Let $u^i = w^T z^i$.

$$\frac{\partial l}{\partial w} = \sum_i 2(y^i - \sigma(u^i))(-1)\sigma(u^i)(1 - \sigma(u^i))z^i$$

Let $s^i = \alpha^T x^i$ and $t^i = \beta^T x^i$.

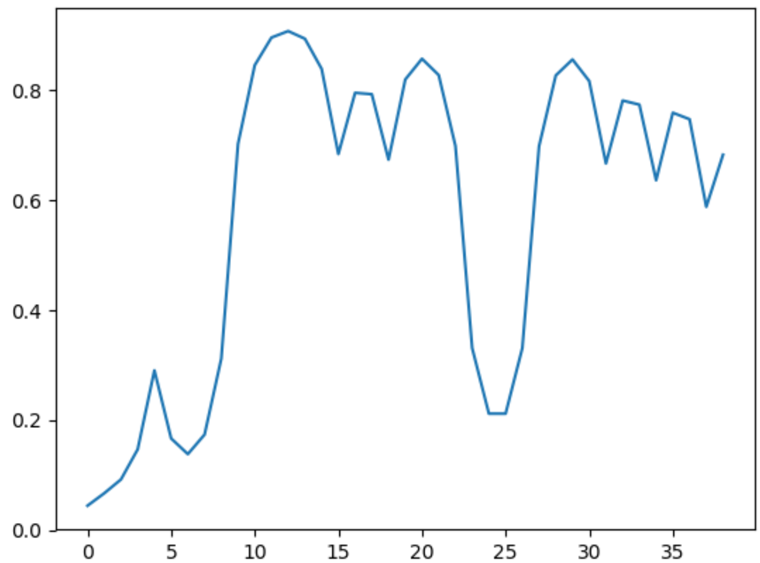
$$\frac{\partial l}{\partial \alpha} = \sum_i 2(y^i - \sigma(u^i))(-1)\sigma(u^i)(1 - \sigma(u^i))w_1\sigma(s^i)(1 - \sigma(s^i))x^i$$

$$\frac{\partial l}{\partial \beta} = \sum_i 2(y^i - \sigma(u^i))(-1)\sigma(u^i)(1 - \sigma(u^i))w_2\sigma(t^i)(1 - \sigma(t^i))x^i$$

5.

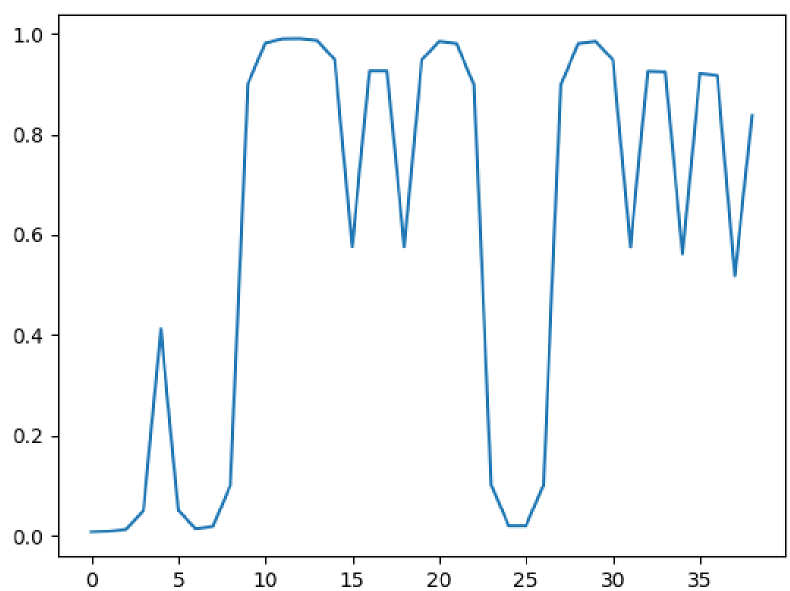
(a)

The probability that the economy is in a good state in the week 39 is 0.683



(b)

The probability that the economy is in a good state in the week 39 is 0.8379



Question 6. Let's find the dual problem of hard margin SVM first.

$$\begin{aligned}&min\ \frac{1}{2}w^Tw\\&s.t.\ 1-y^i(w^Tx^i+b)\leq 0\end{aligned}$$

Lagrangian form is

$$L=\frac{1}{2}w^Tw+\sum_{i=1}^{m}\alpha_i(1-y^i(w^Tx^i+b))$$

$$\frac{\partial L}{\partial w}=w-\sum_{i=1}^{m}\alpha_iy^ix^i=0$$

$$\frac{\partial L}{\partial b}=-\sum_{i=1}^{m}\alpha_iy^i=0$$

Plug back and we get

$$\begin{aligned}L&=\frac{1}{2}(\sum\alpha_iy^ix^i)^T(\sum\alpha_iy^ix^i)+\sum_{i=1}^{m}\alpha_i(1-y^i((\sum\alpha_iy^ix^i)^Tx^i+b))\\&=\sum_i\alpha_i-\frac{1}{2}\sum_{i,j}\alpha_i\alpha_jy^iy^j(x^i)^Tx^j\end{aligned}$$

$\therefore$ the dual problem of hard margin SVM is

$$\begin{aligned}&max_\alpha\ \sum_i\alpha_i-\frac{1}{2}\sum_{i,j}\alpha_i\alpha_jy^iy^j(x^i)^Tx^j\\&s.t.\ \alpha_i\geq 0\ \forall 1\leq i\leq m,\\&\sum_{i=1}^{m}\alpha_iy^i=0\end{aligned}$$

Similarly, the dual problem of soft margin SVM is

$$\begin{aligned}&max_\alpha\ \sum_i\alpha_i-\frac{1}{2}\sum_{i,j}\alpha_i\alpha_jy^iy^j(x^i)^Tx^j\\&s.t.\ C\geq\alpha_i\geq 0\ \forall 1\leq i\leq m,\\&\sum_{i=1}^{m}\alpha_iy^i=0\end{aligned}$$

The only difference between two is there are constraints $C\geq\alpha_i$ in dual problem of soft margin SVM. Suppose solution for dual problem of hard margin SVM is $(\alpha_i^\star,\cdots,\alpha_m^\star)$. If we set C to be $max(\alpha_1^\star,\cdots,\alpha_m^\star)$, then solution for dual problem of soft margin SVM will be the same as that of hard margin SVM.