Assignment3

June 10, 2019

기계학습 숙제 4

1. [SVM; 40p] 선형적으로 분리가 가능한 경우, 다음을 증명하세요.

A. SVM의 원시 형태 (primal form) 해를 구하는 방법
선형적으로 분리가 가능한 이진 분류를 하는 경우 hyperplane으로부터 상대적인 거리를 구하면
$g(x_t) = w^T x_t + b >= +1$ 이면 $y_t = +1$
그리고
$g(x_t) = w^T x_t + b <= -1$ 이면 $y_t = -1$
위의 두 식을 하나로 합치면
$y_t(w^T x_t + b) >= +1$
SVM은 support vector와 hyperplane의 margin을 최대화는 방식이기 때문에
support vector와 hyperplane의 margin는
$margin = |\frac{+1}{||w||} - \frac{-1}{||w||}|$
위의 margindmf 최대화하는 것은 $||w||$를 최소화 하는 것이다.
SVM primal의 해를 구하는 것은
$minimize_w \frac{1}{2}||w||^2$
subject to $y_t(w^T x_t + b) >= +1, \forall t \in [1, N]$
이는 convex한 문제 이므로 아래의 quadratic programming에 의해서
$minimize_z \frac{1}{2} z^T Q z + c^T z$
subject to $Az >= a$
quadratic programming으로 SVM primal을 풀면 다음과 같다.
$z = \begin{bmatrix} b \\ w \end{bmatrix} \in R^{d+1}$
$\frac{1}{2} = \begin{bmatrix} b \\ w^T \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} b \\ w^T \end{bmatrix} = z^T \begin{bmatrix} 0 & 0 \\ 0 & I \end{bmatrix} z$
따라서
$Q = \begin{bmatrix} 0 & 0 \\ 0 & I \end{bmatrix}, c = 0$
$y_n(w^T x_n + b) >= +1 = \begin{bmatrix} y_n & y_n x_n^T \end{bmatrix} z >= 1$
위의 식에서 $\begin{bmatrix} y_n & y_n x_n^T \end{bmatrix}$을 solver에게 전달하면 SVM의 해를 구하게 된다.
B. SVM의 이중 형태 (dual form) 해를 구하는 방법
dual form으로 해를 구하기 위해서는
step 1: compute Lagrangian
primal form이 아래와 같이 주어진다면
$minimize_w \frac{1}{2}||w||^2$
subject to $y_t(w^T x_t + b) >= +1, \forall t \in [1, N]$
Lagrangian function을 사용하여 아래의 식으로 변경한다.

$L(w,b,\alpha) = \frac{1}{2}||w||^2 - \sum_{t=1}^{N} \alpha_t [y_t(w^T x_t + b - 1)]$
where $\alpha >= 0$
위의 식에서 dual form을 구하기 위해 아래의 세개의 KKT conditionds을 만족하게 하면

1. $\frac{\partial L(w^*, b^*, \alpha^*)}{\partial w} = 0$
2. $\frac{\partial L(w^*, b^*, \alpha^*)}{\partial b} = 0$
3. $\alpha_t^* [y_t(w^{*T} x_t + b^* - 1)] = 0, \forall t \in [1, N]$

우선 1번과 2번의 의해서
$\frac{\partial L}{\partial w} = w - \sum_{t=1}^{N} \alpha_t y_t x_t = 0$
$w = \sum_{t=1}^{N} \alpha_t y_t x_t$
$\frac{\partial L}{\partial b} = -\sum_{t=1}^{N} \alpha_t y_t = 0$
$\sum_{t=1}^{N} \alpha_t y_t = 0$
그리고 3번에 의해서 아래와 같은 의미를 내포한다.
if $\alpha > 0$, $\alpha_t^* [y_t(w^{*T} x_t + b^* - 1)] = 1$
if $\alpha_t^* [y_t(w^{*T} x_t + b^* - 1)] > 1$, $\alpha > 0$
그리고 Lagrangian function을 이용하여 구한 식에 KKT 조건 1번과 2번을 이용하여 식을 변경하면 아래와 같은 dual form이 된다.
$maximize_\alpha \sum_{t=1} N\alpha - \frac{1}{2} \sum_{s=1}^{N} \sum_{t=1}^{N} \alpha_s \alpha_t y_s y_t x_s^T x_t$
subject to
$\sum_{t=0}^{N} \alpha_t y_t = 0$
$\alpha_t >= 0, \forall t \in [1, N]$
이제 dual form을 quadaratic programming을 이용하여 풀면
$QP = minmize_x \frac{1}{2} x^T Q x + c^T x$
subject to
$Ax <= b$
$Ex = d$
dual form은 아래와 같이 변한다.
$minmize_\alpha \frac{1}{2} \alpha^T Q \alpha + 1^T \alpha$
subject to
$-\alpha <= 0$
$y^T \alpha = 0$
$Q, 1^T, y^T$를 solver에게 주어 계산하면 된다.

2. [Kernel; 20p] 다음의 함수가 valid kernel임을 보이세요.

A. 임의의 행렬 A 에 대한 K(x,z) = $x^T A^T A z$
$\Phi(x) = Ax$ 라고하면
K(x,z) = $\Phi(x)^T \Phi(z) = (Ax)^T A z = x^T A^T A z$ 이므로 valid kernel이다.
B. K(x,z) = $(x^T z + C)^2$
K(x, z) = $C^2 + Cx_1z_1 + Cx_2z_2 + Cx_1z_1 + x_1^2z_1^2 + x_1z_1x_2z_2 + Cx_2z_2 + x_1z_1x_2z_2 + x_2^2z_2^2 = C^2 + 2Cx_1z_1 + 2Cx_2z_2 + x_1^2z_1^2 + 2x_1z_1x_2z_2 + x_2^2z_2^2 = \Phi(x)^T \Phi(z)$
where
$\Phi(x) = (C, x_1^2, x_2^2, \sqrt{2C}x_1, \sqrt{2C}x_2, \sqrt{2}x_1x_2)$
$\Phi(z)$를 구한다면
$\Phi(z) = (C, z_1^2, z_2^2, \sqrt{2C}z_1, \sqrt{2C}z_2, \sqrt{2}z_1z_2)$
따라서
$\Phi(x)^T \Phi(z)$ = K(x, z)와 같아 valid kernel이다.

3. [HMM; 40p] It is well known that a DNA sequence is a series of components from {A, C, G, T}. Now let's assume there is one hidden variable S that controls the generation of DNA sequence. S takes 2 possible states {S1, S2}. Assume the following transition probabilities for HMM M.

$$P(S1|S1) = 0.8, P(S2|S1) = 0.2, P(S1|S2) = 0.2, P(S2|S2) = 0.8$$

emission probabilities as following

$$P(A|S1) = 0.4, P(C|S1) = 0.1, P(G|S1) = 0.4, P(T|S1) = 0.1$$

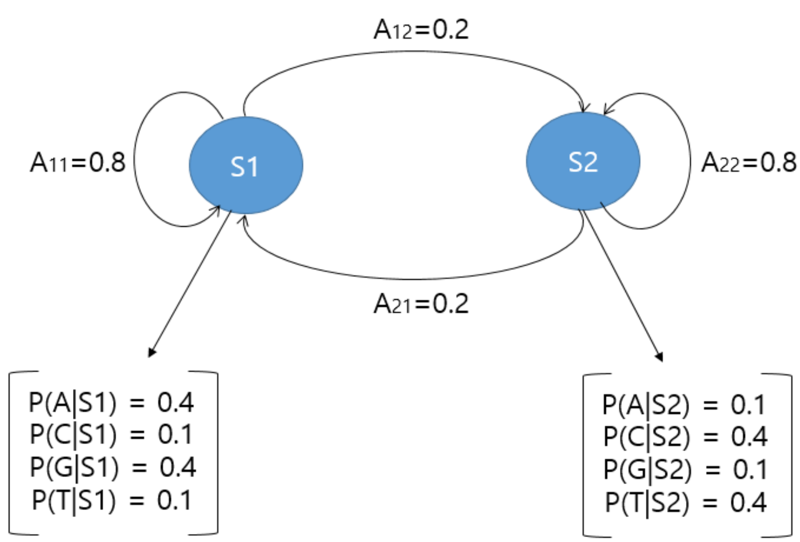
$$P(A|S2) = 0.1, P(C|S2) = 0.4, P(G|S2) = 0.1, P(T|S2) = 0.4$$

and starting probabilities as following

$$P(S1) = 0.5, P(S2) = 0.5$$

Assume the obserbed sequence is x = CGTCAG, calculate:

A. Draw the state diagram of this HMM M and mark the transition probabilities



B. P(x|M) using the forward algorithm. show your work to get full credit.

$$\forall S_k \in \{S_1, S_2\}, \alpha_1^k = P(x_1|\pi_1 = S_k)P(\pi_1 = S_k)$$

$$\forall S_k \in \{S_1, S_2\}, t \in \{2, 3, 4, 5, 6\}$$

$$\alpha_t^k = P(x_t|\pi_t = S_k) \sum_{i \in \{1,2\}} \alpha_{t-1}^i A_{i,k}$$

α	$S_k = S_1$	$S_k = S_2$
α_1^k	0.05	0.2
α_2^k	0.032	0.017
α_3^k	0.0029	0.008
α_4^k	3.9200e-04	0.0028
α_5^k	3.4880e-04	2.3120e-04
α_6^k	1.3011e-04	2.5472e-05

C. The posterior probabilities P(π_i = S1|x, M) for i = 1,...6. show your work to get full credit.

$$\forall S_k \in \{S_1, S_2\}, \beta_6^k = 1$$

$$\forall S_k \in \{S_1, S_2\}, t \in \{1, 2, 3, 4, 5\}$$

$$\beta_t^k = \sum_{j \in \{1,2\}} A_{k,j} P(x_{t+1} | \pi_{t+1} = S_k) \beta_{t+1}^j$$

β	$S_k = S_1$	$S_k = S_2$
β_1^k	7.9744e-04	5.7856e-04
β_2^k	0.0022	0.0051
β_3^k	0.0122	0.0150
β_4^k	0.1120	0.0400
β_5^k	0.3400	0.1600
β_6^k	1	1

$$P(\pi_i = S1 | x, M) = \frac{\alpha_i^1 \beta_i^1}{\sum_{k \in \{1,2\}} \alpha_i^k \beta_i^k}$$

$P(\pi_i = S1 \vert x, M)$	$S_k = S_1$
$P(\pi_1 = S1 \vert x, M)$	0.2563
$P(\pi_2 = S1 \vert x, M)$	0.4476
$P(\pi_3 = S1 \vert x, M)$	0.4476
$P(\pi_4 = S1 \vert x, M)$	0.2822
$P(\pi_5 = S1 \vert x, M)$	0.7622
$P(\pi_6 = S1 \vert x, M)$	0.8363

D. The most likely path of hidden states using the Viterbi Algorithm. Show your work to get full credit.

$$b_t^j = argmax_{i=1}^n [v_{t-1} * A_{ij} * P(x_t | \pi_t = S_j)]$$

V_i^k	$S_k = S_1$	$S_k = S_2$
V_i^k	0.0500	0.2000
V_i^k	0.0160	0.0160
V_i^k	0.0013	0.0013
V_i^k	1.0240e-04	0.016
V_i^k	1.3107e-04	1.3107e-04
V_i^k	4.1943e-05	1.0486e-05

thus, the most likely path is S2, S2, S2, S2, S1, S1