תרגיל בית 3– סיווג בייסיאני ורגרסיה

תאריך הגשה: 28.05.2019

מוריאל בן משה **304832512**, דניאל אנגלסמן **300546173**

חלק 1: מסווג בייסיאני

שאלה מס' 1: חימום

נתון כי

$$P(x \mid Y=0) = \mathcal{N}(0,1)$$
$$P(x \mid Y=1) = \mathcal{N}(1,0.5)$$
$$P(Y=1) = 0.4$$

כאשר Y הוא מצב העולם.

1. ללא מדידה כלשהי, מהו המשערך הבייאסיאני האופטימלי?
2. בהינתן מדידה בודדת x , ציירו את ההתפלגויות $P(Y=0,x), P(Y=1,x)$ וחשבו את כלל ההחלטה לסיווג מצב העולם Y.
3. בהינתן זוג מדידות x_1, x_2 , חשבו את כלל ההחלטה לסיווג מצב העולם Y.

1

$$\text{No measurement means that optimal Bayesian estimator equals } Y=0 \tag{1.1}$$

The general term for the estimator :

$$f^*(x) = f_{MAP}(x) \triangleq arg \max_{C \in Y} P_{Y|X}(C|x) \quad \Rightarrow_{Bayes} \quad arg \max_{C \in Y} P_{X|Y}(x|C)P_Y(C)$$
$$P_{Y|X}(x|C_k) = \frac{1}{(2\pi)^{d/2}|\Sigma_k|} exp\Big(-\frac{1}{2}(x-\mu)^T \Sigma_k^{-1}(x-\mu)\Big) P_Y(C_k)$$

Using the joint pdf relation - $p(x,y) = P(X|Y) \cdot P(Y)$, we'll write :

$$P(x|Y=0) = \mathcal{N}(0,1) \cdot (1 - P(Y=1)) = \frac{1}{\sqrt{2\pi}} exp\Big(-\frac{1}{2}x^T x\Big) \cdot 0.6$$
$$P(x|Y=1) = \mathcal{N}(1,0.5) \cdot P(Y=1) = \frac{1}{\sqrt{2\pi|0.5|}} exp\Big(-\frac{1}{2}(x-1)^T 0.5^{-1}(x-1)\Big) \cdot 0.4$$

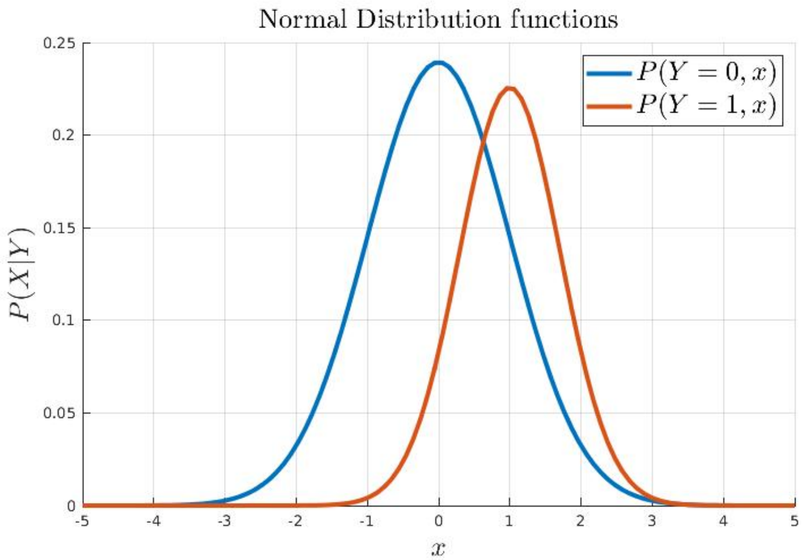

Figure 1: Distribution of both conditional probabilities

Define the minimized terms (after assigning log expression) :

$$P(x|Y=0) = ln(\frac{0.6}{\sqrt{2\pi}}) - \frac{1}{2}x^T x = -1.429 - 0.5x^2$$
$$P(x|Y=1) = ln(\frac{0.4}{\sqrt{2\pi|0.5|}}) - \frac{1}{2}(x-1)^T 0.5^{-1}(x-1) = -1.488 - (x-1)^2$$
$$g_{kj}(x) \triangleq 0.5x^2 - 2x + 1.059 \geq 0 \quad \rightarrow \quad x < 0.63 \text{ or } 3.37 < x \qquad (1.2)$$

For the bi-variate case ($\bar{x} = [x_1,\ x_2]^T$), we get :

$$P(x_1, x_2|Y=0) = ln(\frac{0.6}{2\pi}) - \frac{1}{2}(x_1^2 + x_2^2) = -1.429 - \frac{x_1^2 + x_2^2}{2}$$
$$P(x|Y=1) = ln(\frac{0.4}{2\pi|0.5|}) - (\bar{x}-1)^T(\bar{x}-1) = -1.488 - \Big((x_1-1)^2 + (x_2-1)^2\Big)$$
$$g_{kj}(\bar{x}) \triangleq (\mathbf{x}_1 - 1)^2 + (\mathbf{x}_2 - 1)^2 - \frac{1}{2}(x_1^2 + x_2^2) - 0.287 \geq 0 \quad (1.3)$$

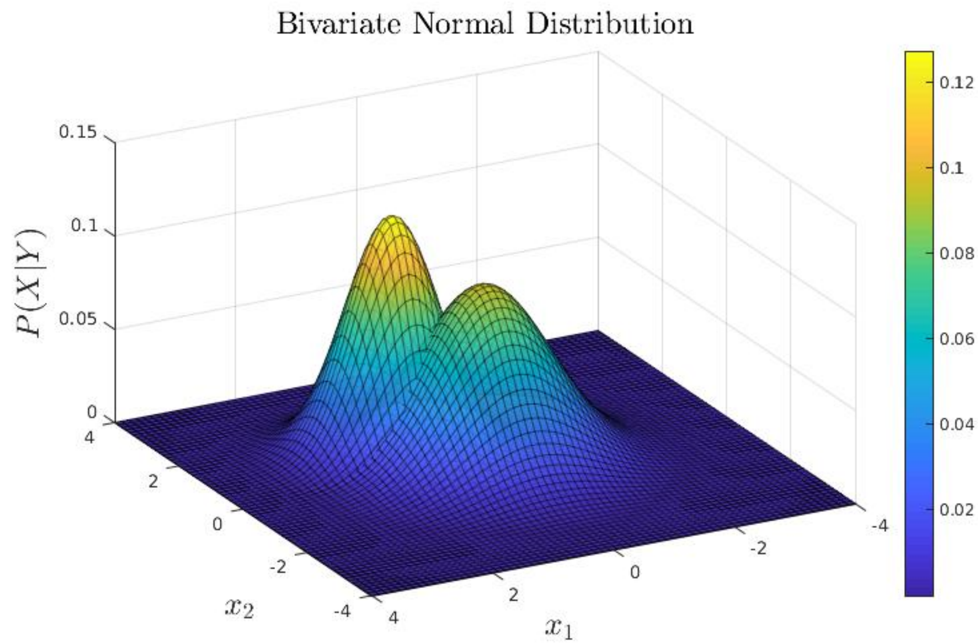

Figure 2: Bi-variate distribution with an elliptic decision place $(x_r, y_r) = (1, 1)$

2

נרצה להשתמש במודל Naïve Bayes אשר מסווג את המסמך על סמך וקטור השכיחות $\vec{m}$ והמודל ההסתברותי שהוגדר לעיל. הניחו כי כל הקבועים שצוינו ידועים, אלא אם נאמר אחרת. פונקציית הסיכון הינה הסתברות השגיאה הרגילה (הפסד 0-1).

א. (1) חשבו את הקבועים A_j, B_j (במונחים של α_j, β_j).

(2) רישמו את ההסתברות שמסמך בעל וקטור-מילים $\vec{m}$ הינו ספאם (בהנחות המתאימות למודל האמור).

ב. הוכיחו שכלל ההחלטה האופטימאלי המתקבל תחת הנחת Naïve Bayes הינו לינארי, מהצורה $\hat{y} = \mathrm{sign}\left(w^T \vec{m} + b\right)$. רישמו במפורש את w, b.

ג. בסעיף זה (בלבד) נדרש להעריך את הקבועים α_j בעזרת דוגמאות. הניחו כי נתונים N מסמכי ספאם, בעלי וקטורי שכיחות $\{\vec{m}(i)\}_{i=1}^N$. חשבו את משערך הסבירות המרבית (MLE) של כל פרמטר α_j.

2.1

We'll denote the relative likelihood of the data as :

$$\hat{p}(y=-1)_{spam} = \frac{n_k}{n} = p_0, \qquad \hat{p}(y=1)_{docs} = \frac{n-n_k}{n} = 1-p_0$$

The j index denotes a certain word out of **J** *different* words, while m is the likelihood :

$$\sum_{m=1}^{J} P(m_j = m|y=-1) = \sum_{m=1}^{J} A_j \cdot (\alpha_j)^m = 1 \quad \rightarrow \quad A_j \frac{1}{1-\alpha_j} = 1$$
$$\text{therefore,} \qquad A_j = 1-\alpha_j, \qquad B_j = 1-\beta_j \tag{2.1}$$

2.2

Given an a-posteriori conditional : $$P(y=-1|\bar{m}) = \frac{P(\bar{m}|y=-1)P(y=-1)}{P(\bar{m})}$$

$$P(y=-1|\bar{m}) = \frac{\prod_{j=1}^{J} A_j(\alpha_j)^{m_j} p_0}{\prod_{j=1}^{J} A_j(\alpha_j)^{m_j} p_0 + \prod_{j=1}^{J} B_j(\beta_j)^{m_j}(1-p_0)}$$

2.3

Calculate the decision plane :

$$
\begin{aligned}
f_{MAP}(x) = arg \max_{C \in Y} P_{Y|X}(x|C_k) &= arg \max_{C \in Y} P_{X|Y}(X|Y)P(Y) \\
\Rightarrow \quad & P_{X|Y}(x|C_k)P_Y(C_k) \geq P_{X|Y}(x|C_j)P_Y(C_j) \\
\text{here,} \qquad \prod_{j=1}^{J} A_j(\alpha_j)^{m_j} p_0 &\geq \prod_{j=1}^{J} B_j(\beta_j)^{m_j}(1-p_0) \qquad \backslash \cdot \log(\ ) \\
log(p_0) + \sum_{j=1}^{J} log\big(A_j(\alpha_j)^{m_j}\big) &\geq log(1-p_0) + \sum_{j=1}^{J} log\big(B_j(\beta_j)^{m_j}\big) \\
log(\frac{1}{p_0}) - \sum_{j=1}^{J} logA_j - \sum_{j=1}^{J} m_j log(\alpha_j) + \sum_{j=1}^{J} logB_j &+ \sum_{j=1}^{J} m_j log(\beta_j) \leq 0 \\
log(\frac{1}{p_0}) + \sum_{j=1}^{J} log(\frac{B_j}{A_j}) &+ \sum_{j=1}^{J} m_j \cdot log(\frac{\beta_j}{\alpha_j}) \leq 0
\end{aligned}
$$

Extract the equivalent elements to the linear decision plane :

$$
\hat{y} = \text{sign } (w^T \bar{m} + b) = \text{sign } \Big(log(\frac{\beta_j}{\alpha_j}) \cdot \bar{m} + \big(log(\frac{1}{p_0}) + \sum_{j=1}^{J} log(\frac{B_j}{A_j}) \big) \Big) \tag{2.2}
$$

2.4

Willing to maximize the log-likelihood function of N spam files :

$$
\begin{aligned}
log\big(L(\alpha_j, \{\bar{m}(i)\}_{i=1}^{N}\big) = \sum_{i=1}^{N} log(1-\alpha_j) + \sum_{i=1}^{N} m_i^j \cdot log(\alpha_j) &= Nlog(1-\alpha_j) + log(\alpha_j) \sum_{i=1}^{N} m_i^j \\
\frac{d}{d\alpha_j}\ log\big(L(\alpha_j, \{\bar{m}(i)\}_{i=1}^{N}\big) &= -\frac{N}{1-\alpha_j} + \frac{1}{\alpha_j} \sum_{i=1}^{N} m_i^j = 0
\end{aligned}
$$

Finlally,

$$
\frac{1-\alpha_j}{\alpha_j} = \frac{N}{\sum_{i=1}^{N} m_i^j} \quad \Rightarrow \quad \frac{1}{\alpha_j} = \frac{N}{\sum_{i=1}^{N} m_i^j} + 1 \quad \Rightarrow \quad \alpha_j = \underline{\underline{\frac{\sum_{i=1}^{N} m_i^j}{\sum_{i=1}^{N} m_i^j + N}}} \tag{2.3}
$$

3

3.1

$$Relate \quad expressions: \tag{3.1}$$

- ○ i = c ($\lambda = 1$) - light restraint on w_1 allows big slope.
- ○ ii = b ($\lambda = 10$) - heavy restraint on w_1 allows small slope.
- ○ iii = a ($\lambda = 1$) - both weights are slightly restrained.
- ○ iv = d ($\lambda = 10$) - both weights are heavily restrained.

ב. נגדיר את $(\tilde{w}_1^*, \tilde{w}_0^*)$ להיות פתרון בעיית הרגרסיה a בסעיף א'. חשבו את ערך הביטוי $\sum_{i=1}^{3}(y_i - \tilde{w}_1^* x_i - \tilde{w}_0^*)$

ג. מצאו את ערך פונקציית המטרה בנקודה האופטימלית (w_1^*, w_0^*) עבור בעיית הרגרסיה הבאה: $\sum_{i=1}^{3}(y_i - w_1 x_i - w_0)^2$, עבור קבוצת האימון מסעיף א.

ד. כעת נגדיר את בעיית הרגרסיה הבאה: $\sum_{i=1}^{3}(y_i - w_1 x_i - w_0)^2 + \lambda \cdot \mathbb{I}_{[w_1 \neq 0]}$, כאשר $\mathbb{I}_{[w_1 \neq 0]} = \begin{cases} 1 & \text{if } w_1 \neq 0 \\ 0 & \text{else} \end{cases}$

a. חשבו את פתרון בעיית הרגרסיה (w_1^*, w_0^*), וערך פונקציית המטרה כפונקציה של הקבוע λ

b. מהו פתרון בעיית הרגרסיה (w_1^*, w_0^*), וערך פונקציית המטרה בפתרון זה כאשר $\lambda \to \infty$?

3.2

Let $\tilde{w}_1^*, \tilde{w}_2^*$ be the solution of the regression :

$$\sum_{i=1}^{3}(y_i - \tilde{w}_1^* x_i \tilde{w}_0^*) = 4 - 3\tilde{w}_0^* - 3\tilde{w}_1^* \tag{3.2}$$

3.3

Given a 2nd order problem a which is twice differentiable, the problem is said to be convex if $\nabla^2 f > 0$. Here, both weights are restrained by $\lambda = 1$, and its Hessian fulfills :

$$\mathbf{H}_{i,j} = \frac{\partial^2 f}{\partial w_{0,i} \partial w_{1,j}} = 0, \quad \forall \quad i \neq j$$

The general case for finding optimal w^* with regularization :

$$\begin{aligned}
E(w) = \|Y - \Phi w\| &= \sum_{i=1}^{n} (y_i - \hat{f}_w(x))^2 + \lambda w^T w \\
\mathbf{X} = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}, \quad \mathbf{Y} &= \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}, \quad \Phi = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \\
Q = \Phi^T \Phi = \begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix}, \quad B = \Phi^T \mathbf{Y} &= \begin{bmatrix} 4 \\ 5 \end{bmatrix} \quad c = \mathbf{Y}^T \mathbf{Y} \\
E(w) = w^T Q w - 2w^T B + c + \lambda w^T w \quad &\rightarrow \quad \frac{d}{dw} E(w) = 0 \\
2Qw - 2B + 2\lambda w = 0 \quad &\rightarrow \quad w^* = (Q + \lambda I)^{-1} B \\
\text{No regularization } (\lambda = 0), \quad &w^* = [\frac{5}{6}, \ \frac{1}{2}]^T
\end{aligned} \tag{3.3}$$

Plug inside equation, and obtain the general **loss** :

$$\hat{L}_n(f_w, S) = E(w^*, \lambda) = \big(\frac{35}{6} - 2 \cdot \frac{35}{6} + 6 + \frac{17}{18}\lambda\big) = \frac{1}{6} + \frac{17}{18}\lambda$$

3.4

Examine both λ's, but first check the case of $\bar{w}(w_1 = 0) = \begin{bmatrix} w_0 \\ 0 \end{bmatrix}$:

$$\begin{aligned}
E(w_1 = 0) = w^T Q w - 2w^T B + c + \lambda w^T w &= (3 + \lambda) w_0^2 - 8w_0 + \mathbb{I}_\lambda \cdot \lambda w_0^2 \\
\frac{d}{dw} E(w) = 2(3 + \lambda) w_0 - 8 = 0 \quad &\Rightarrow \quad \bar{w}^* = \begin{bmatrix} w_0^* \\ w_1^* \end{bmatrix} = \begin{bmatrix} \frac{4}{3+\lambda} \\ 0 \end{bmatrix}
\end{aligned}$$

$$(a) \quad \mathbb{I}(\lambda_{const}) = \begin{cases} 1, & \text{if } w_1 \neq 0 \;\; \rightarrow \qquad \hat{L}_n([w_0, w_1],\ S) = \frac{1}{6} + \frac{17}{18}\lambda \\ 0, & \text{else} \quad\;\; \rightarrow_{w_1=0} \quad \hat{L}_n([w_0, 0],\ S) = \frac{2}{3} \end{cases}$$

$$(b) \quad \mathbb{I}(\lambda \rightarrow \infty) = \begin{cases} 1, & \text{if } w_1 \neq 0 \;\; \rightarrow \qquad \hat{L}_n([0, w_1],\ S) = \frac{2}{3} \\ 0, & \text{else} \quad\;\; \rightarrow_{w_1=0} \quad \hat{L}_n([0, 0],\ S) = \frac{2}{3} \end{cases}$$

3.5

נגדיר את שני הגדלים הבאים עבור סט נקודות כללי $\{x_i, y_i\}_{i=1}^n$ עבור $x_i, y_i \in \Re$:

$$C_{xy} = \frac{1}{n}\sum_{t=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) \quad , \quad C_{xx} = \frac{1}{n}\sum_{t=1}^{n}(x_i - \overline{x})^2$$

כאשר $\overline{x} = \frac{1}{n}\sum_{t=1}^{n} x_i, \ \overline{y} = \frac{1}{n}\sum_{t=1}^{n} y_i$. בטאו את w_1^* בפתרון לבעיית הרגרסיה $\sum_{i=1}^{n}(y_i - w_1 x_i - w_0)^2$ באמצעות C_{xy}, C_{xx}.

$$\hat{L}_n = \frac{1}{n}\sum_{i=1}^{n}(y_i - w_1 x_i - w_0)^2 = \frac{1}{n}\sum_{i=1}^{n}\Big(y_i^2 - 2y_i w_1 x_i - 2y_i w_0 + (w_1 x_i)^2 + 2w_1 x_i w_0 + (w_0)^2\Big)$$

$$\frac{\partial}{\partial w_0}(\hat{L}_n) = \frac{2}{n}\sum_{i=1}^{n}\Big(w_0^* + w_1^* x_i - y_i\Big) = 0 \quad \Rightarrow \quad w_0^* = \frac{1}{n}\sum_{i=1}^{n}\Big(y_i - w_1^* x_i\Big) = \bar{y} - w_1^*\bar{x}$$

$$\frac{\partial}{\partial w_1}(\hat{L}_n) = \frac{2}{n}\sum_{i=1}^{n} x_i\Big((w_0^* - y_i) + w_1^* x_i\Big) = 0$$

$$\text{Plug inside :} \quad \frac{\partial}{\partial w_1}(\hat{L}_n) = \frac{2}{n}\sum_{i=1}^{n} x_i\Big((\bar{y} - w_1^*\bar{x} - y_i) + w_1^* x_i\Big) = 0$$

$$\frac{2}{n}\sum_{i=1}^{n} x_i\Big((\bar{y} - y_i) + w_1^*(x_i - \bar{x})\Big) = 0 \quad \Rightarrow \quad w_1^* \cdot \frac{2}{n}\sum_{i=1}^{n}(x_i - \bar{x}) = \frac{2}{n}\sum_{i=1}^{n}(y_i - \bar{y}) \quad \backslash \cdot (x_i - \bar{x})$$

$$w_1^* \cdot \frac{2}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{2}{n}\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})$$

$$w_1^* C_{xx} = C_{xy} \quad \Rightarrow \quad \underline{\underline{w_1^* = \frac{C_{xy}}{C_{xx}}}} \tag{3.4}$$

4

בשאלה זו נרצה לחקור את בעיית הרגרסיה עם רגולריזציה. נתונות לנו N נקודות:

$\{\mathbf{x}_n, y_n\}_{n=1}^{N}$ כאשר $\mathbf{x}_n \in \mathbb{R}^d, y_n \in \mathbb{R}$. פונקציית המחיר מוגדרת באופן הבא:

$$L(\mathbf{w}) = \frac{1}{2}\sum_{i=1}^{N}[(\mathbf{w}^{\mathbf{T}}\mathbf{x}_n - y_n)^2] + \frac{\lambda}{2}\|\mathbf{w}\|^2, \qquad \lambda > 0$$

א. נגדיר $\mathbf{y} = (y_1 \quad \cdots \quad y_n) \in \mathbb{R}^n$, $\mathbf{X} = (\mathbf{x}_1 \quad \cdots \quad \mathbf{x}_n) \in \mathbb{R}^{d\times n}$. בטאו את פונקציית המחיר באמצעות $(\mathbf{X}, \mathbf{y}, \mathbf{w})$.

ב. מצאו את וקטור המשקולות $\mathbf{w}$ הממזער את פונקציית המחיר והביעו אותו באמצעות $(\mathbf{X}, \mathbf{w})$.

4.1

Let $\mathbf{w} \in \mathbb{R}^d$:

$$\text{Scalar}: \quad L(\mathbf{w}) = \frac{1}{2}\sum_{i=1}^{N}[(\mathbf{w}^T\mathbf{x}_i - y_i)^2] + \frac{\lambda}{2}\|\mathbf{w}\| \in \mathbb{R}$$

$$\text{Vectorical}: \quad L(\mathbf{w}) = \frac{1}{2}[\mathbf{X}^T\mathbf{w} - \mathbf{y}]^T[\mathbf{X}^T\mathbf{w} - \mathbf{y}] + \frac{\lambda}{2}\mathbf{w}^T\mathbf{w}$$

$$L(\mathbf{w}) = \frac{1}{2}[\mathbf{w}^T\mathbf{X}\mathbf{X}^T\mathbf{w} - \mathbf{w}^T\mathbf{X}\mathbf{y} - \mathbf{y}^T\mathbf{X}^T\mathbf{w} + \mathbf{y}^T\mathbf{y}] + \frac{\lambda}{2}\mathbf{w}^T\mathbf{w} \tag{4.1}$$

4.2

Express $\mathbf{w}^*$ that minimizes the cost function :

$$\frac{d}{dw}L(\mathbf{w}) = \mathbf{X}\mathbf{X}^T\mathbf{w} - \mathbf{X}\mathbf{y} + \mathbf{w}\lambda = 0, \quad \in \mathbb{R}^d$$

$$\left(\mathbf{X}\mathbf{X}^T + \lambda\mathbf{I}\right)\mathbf{w} = \mathbf{X}\mathbf{y}$$

$$\mathbf{w}^* = \left(\mathbf{X}\mathbf{X}^T + \lambda\mathbf{I}\right)^{-1}\mathbf{X}\mathbf{y} \in \mathbb{R}^d \tag{4.2}$$

Assuming $\left(\mathbf{X}\mathbf{X}^T + \lambda\mathbf{I}\right)^{-1}$ is positive-definite and thus inverse.

ג. מצאו את וקטור המשקולות $\mathbf{w}$ הממזער את פונקציית המחיר. הביעו את וקטור המשקולות באמצעות $(\mathbf{\Phi}, \mathbf{y}, \mathbf{w})$. מהו המימד של הוקטור $\mathbf{w}$ המתקבל?

4.3

Let the transformation $\phi(\cdot) : \mathbb{R}^d \to \mathbb{R}^m$, such that $\mathbf{A}$ is the transformation matrix :

$$\begin{aligned}
\phi(x_i) = \mathbf{A}\mathbf{x}_i \ \in \mathbb{R}^m \ , \quad \text{given :} \quad \mathbf{A} \in \mathbb{R}^{m\times d}, \quad \mathbf{X} \in \mathbb{R}^{d\times n} \\
\Rightarrow \quad \Phi = \mathbf{A}\big(\mathbf{x}_1, \ldots, \mathbf{x}_n\big) = \mathbf{A}\mathbf{X} \ \in \mathbb{R}^{m\times n} \ , \quad \mathbf{w} \ \in \ \mathbb{R}^m \\
\text{Scalar :} \quad L(\mathbf{w}) = \frac{1}{2}\sum_{i=1}^{N}[(\mathbf{w}^T\phi(x_i) - y_i)^2] + \frac{\lambda}{2}\|\mathbf{w}\| \ \in \ \mathbb{R} \\
L(\mathbf{w}) = \frac{1}{2}\sum_{i=1}^{N}[(\mathbf{w}^T\mathbf{A}x_i - y_i)^2] + \frac{\lambda}{2}\|\mathbf{w}\| \ \in \ \mathbb{R}
\end{aligned}$$

Vectorical :

$$\begin{aligned}
L(\mathbf{w}) = \frac{1}{2}[\mathbf{X}^T\mathbf{A}^T\mathbf{w} - \mathbf{y}]^T[\mathbf{X}^T\mathbf{A}^T\mathbf{w} - \mathbf{y}] + \frac{\lambda}{2}\mathbf{w}^T\mathbf{w} \\
L(\mathbf{w}) = \frac{1}{2}[\mathbf{w}^T(\mathbf{A}\mathbf{X})(\mathbf{X}^T\mathbf{A}^T)\mathbf{w} - \mathbf{w}^T(\mathbf{A}\mathbf{X})\mathbf{y} - \mathbf{y}^T(\mathbf{X}^T\mathbf{A}^T)\mathbf{w} + \mathbf{y}^T\mathbf{y}] + \frac{\lambda}{2}\mathbf{w}^T\mathbf{w} \\
L(\mathbf{w}) = \frac{1}{2}[\mathbf{w}^T\Phi\Phi^T\mathbf{w} - \mathbf{w}^T\Phi\mathbf{y} - \mathbf{y}^T\Phi^T\mathbf{w} + \mathbf{y}^T\mathbf{y}] + \frac{\lambda}{2}\mathbf{w}^T\mathbf{w} \\
\frac{d}{dw}L(\mathbf{w}) = \Phi\Phi^T\mathbf{w} - \Phi\mathbf{y} + \mathbf{w}\lambda = 0, \quad \in \ \mathbb{R}^m \\
\big(\Phi\Phi^T + \lambda\mathbf{I}\big)\mathbf{w} = \Phi\mathbf{y} \quad \Rightarrow \quad \mathbf{w}^* = \big(\Phi\Phi^T + \lambda\mathbf{I}\big)^{-1}\Phi\mathbf{y}, \quad \in \ \mathbb{R}^m \qquad (4.3)
\end{aligned}$$

4.4

הניחו כי נתונה לכם פונקציית גרעין שמקיימת $K(\mathbf{x},\mathbf{z}) = \boldsymbol{\phi}(\mathbf{x})^T\boldsymbol{\phi}(\mathbf{z})$.

1) הוכיחו את הזהות הבאה $(\mathbf{BA}+\lambda\mathbf{I})^{-1}\mathbf{B} = \mathbf{B}(\mathbf{AB}+\lambda\mathbf{I})^{-1}$.
2) היעזרו בזהות שהוכחתם והראו כי ניתן להשתמש בטריק הגרעין על מנת לחשב את המשערך בצורה יעילה. כלומר, הראו כי ניתן לחשב את המשערך לכל נקודה ללא חישוב מפורש של $\boldsymbol{\phi}(\mathbf{x}_n)$ ו-$\boldsymbol{\phi}(\mathbf{x})$.

By playing with the inverse property we'll get :

$$\begin{aligned}(\mathbf{BA}+\lambda\mathbf{I})^{-1}\mathbf{B} = \left(\mathbf{B}^{-1}(\mathbf{BA}+\lambda\mathbf{I})\right)^{-1} = (\mathbf{A}+\lambda\mathbf{B}^{-1})^{-1} \\ \left((\mathbf{AB}+\lambda)\mathbf{B}^{-1}\right)^{-1} = \underline{\underline{\mathbf{B}(\mathbf{AB}+\lambda\mathbf{I})^{-1}}}\end{aligned} \tag{4.4}$$

Using the following identity :

$$\begin{aligned}\hat{y}(\bar{x}) = w^T\phi(x) = (w^*)^T\phi(x) = \mathbf{y}^T\Phi^T\left(\Phi\Phi^T+\lambda\mathbf{I}\right)^{-1}\phi(x) \quad \| \quad \mathbf{B}(\mathbf{AB}+\lambda)^{-1} \\ \Leftrightarrow \quad (\mathbf{BA}+\lambda\mathbf{I})^{-1}\mathbf{B} \quad \| \quad \mathbf{y}^T\left(\Phi^T\Phi+\lambda\mathbf{I}\right)^{-1}\Phi\ \phi(x)\end{aligned} \tag{4.5}$$

That way we get :

$$\Phi^T\Phi = \begin{bmatrix} K(x_1,x_1) & \dots & K(x_1,x_n) \\ \vdots & \ddots & \vdots \\ K(x_n,x_1) & \dots & K(x_n,x_n)\end{bmatrix} = K(\Phi^T\Phi)$$

and similarly,

$$\Phi^T\phi(x) = \ldots = K(\Phi^T\phi(x))$$

And we showed the usage of the kernel trick.

$-fin-$