- Show that, irrespective of the dimensionality of the data space, a data set consisting of just two data points (call them $\mathbf{x}^{(1)}$ and $\mathbf{x}^{(2)}$, one from each class) is sufficient to determine the maximum-margin hyperplane. Fully explain your answer, including giving an explicit formula for the solution to the hard margin SVM (i.e., $\mathbf{w}$) as a function of $\mathbf{x}^{(1)}$ and $\mathbf{x}^{(2)}$.

解答：证明给定样本$x^{(1)},x^{(2)}$,能求出线性 SVM 分类器的参数即可。将

$$
\begin{aligned}
&\max_{\alpha} \quad \sum_{i=1}^{n}\alpha_i-\frac{1}{2}\sum_{i,j=1}^{n}y^{(i)}y^{(j)}\alpha_i\alpha_j(x^{(i)})^T x^{(j)}\\
&s.t. \quad \alpha_i\geq 0, i=1,\cdots,n,\\
&\qquad \sum_{i=1}^{n}\alpha_i y^{(i)}=0.
\end{aligned}
$$

将样本代入可得

$$
\max_{\alpha}\left(\alpha_1+\alpha_2-\frac{1}{2}(\alpha_1)^2\left\|x^{(1)}\right\|_2^2+\alpha_1\alpha_2\left(x^{(1)}\right)^T x^{(2)}-\frac{1}{2}(\alpha_2)^2\left\|x^{(2)}\right\|_2^2\right)
$$
$$
\text{s.t.}\quad \alpha_1>0,\alpha_2>0,\qquad \alpha_1-\alpha_2=0
$$

将$\alpha_1=\alpha_2$代入

$$
L(\alpha_1)=\max_{\alpha_1}\left(2\alpha_1-\frac{1}{2}(\alpha_1)^2\left\|x^{(1)}\right\|_2^2+(\alpha_1)^2\left(x^{(1)}\right)^T x^{(2)}-\frac{1}{2}(\alpha_1)^2\left\|x^{(2)}\right\|_2^2\right)
$$

令 $\frac{\partial L(\alpha_1)}{\partial\alpha_1}=2-\alpha_1\left\|x^{(1)}\right\|_2^2+2\alpha_1\left(x^{(1)}\right)^T x^{(2)}-\alpha_1\left\|x^{(2)}\right\|_2^2=0$，可得

$$
\alpha_1=\frac{2}{\left\|x^{(1)}\right\|_2^2-2(x^{(1)})^T x^{(2)}+\left\|x^{(2)}\right\|_2^2}=\frac{2}{\left\|x^{(1)}-x^{(2)}\right\|_2^2}
$$

$$
\alpha_2=\alpha_1=\frac{2}{\left\|x^{(1)}-x^{(2)}\right\|_2^2}
$$

$$
w^*=\alpha_1x^{(1)}-\alpha_2x^{(2)}=\frac{2\left(x^{(1)}-x^{(2)}\right)}{\left\|x^{(1)}-x^{(2)}\right\|_2^2},
$$

$$
\begin{aligned}
b&=1-\left(\alpha_1\left(x^{(1)}\right)^T x^{(1)}-\alpha_2\left(x^{(2)}\right)^T x^{(1)}\right)\\
&=1-\frac{2}{\left\|x^{(1)}-x^{(2)}\right\|_2^2}\left(\left(x^{(1)}\right)^T x^{(1)}-\left(x^{(2)}\right)^T x^{(1)}\right)
\end{aligned}
$$

- Gaussian kernel takes the form:

$$k(\boldsymbol{x}, \boldsymbol{x}') = \exp\left(-\frac{\|\boldsymbol{x} - \boldsymbol{x}'\|^2}{2\sigma^2}\right)$$

  Try to show that the Gaussian kernel can be expressed as the inner product of an infinite-dimensional feature vector.

- **Hint:** Making use of the following expansion, and then expanding the middle factor as a power series.

$$k(\boldsymbol{x}, \boldsymbol{x}') = \exp\left(-\frac{\boldsymbol{x}^T\boldsymbol{x}}{2\sigma^2}\right)\exp\left(\frac{\boldsymbol{x}^T\boldsymbol{x}'}{\sigma^2}\right)\exp\left(-\frac{(\boldsymbol{x}')^T\boldsymbol{x}'}{2\sigma^2}\right)$$

解答：$k(x,x') = \exp(-\frac{x^T x}{2\sigma^2})\exp(\frac{x^T x\prime}{\sigma^2})\exp(-\frac{(x')^T x\prime}{2\sigma^2})$

将中间项$\exp(\frac{x^T x\prime}{\sigma^2})$用泰勒级数展开：

$$\exp\left(\frac{x^T x'}{\sigma^2}\right) = \exp\left(\frac{\sum_{i=1}^{d} x_i x_i'}{\sigma^2}\right) = \sum_{n=0}^{\infty}\frac{\left(\frac{x^T x'}{\sigma^2}\right)^n}{n!}$$

$$= （1，\frac{1}{1!}，\frac{1}{2!},\frac{1}{3!},\frac{1}{4!},\frac{1}{5!},\dots\dots）$$

$$(\left(\frac{x^T x'}{\sigma^2}\right)^0,\left(\frac{x^T x'}{\sigma^2}\right)^1,\left(\frac{x^T x'}{\sigma^2}\right)^2,\dots\dots)^T$$

$$k(x,x') = \left[\exp(-\frac{x^T x}{2\sigma^2})（1，\frac{1}{1!}，\frac{1}{2!},\frac{1}{3!},\frac{1}{4!},\frac{1}{5!},\dots\dots）\right]$$

$$\left[\exp(-\frac{(x')^T x'}{2\sigma^2})\left(\left(\frac{x^T x'}{\sigma^2}\right)^0,\left(\frac{x^T x'}{\sigma^2}\right)^1,\left(\frac{x^T x'}{\sigma^2}\right)^2,\dots\dots\right)\right]^T$$