1.

(1). SVM linear function

$h(\mathbf{x})=sign(\mathbf{w}\text{^}T\ \mathbf{x}+b)$

(2). We have an input vector $\mathbf{x_i}$ (instance)

If $\mathbf{w}^{\mathrm{T}}\mathbf{x_i}+b\geq 0$, $\mathbf{x_i}$ would be classified as positive class; Otherwise, it should be classified as negative class.

2.

(1) Infinite number of decision boundaries if linearly separable.

(2) SVM calculate the boundary with maximum margin.

(3) Model with maximum margin can distinguish samples with better confidence. Also, it normally has better generalization performance as well as better robustness against noise or perturbation.

3.

3.1

(1) Upper bound

The optimization problem can be shown below:

$$\max_{p_i} I(P)=-\left(p_1\log_2 p_1+p_2\log_2 p_2+\cdots+p_n\log_2 p_n\right)$$

$$\text{s.t.}\sum_{i=1}^{n}p_i-1=0$$

$$0\leq p_i\leq 1$$

Define the Lagrange function and calculate the extreme value using Lagrange multiplier:

$$L(P,\beta)=-\left(p_1\log_2 p_1+p_2\log_2 p_2+\cdots+p_n\log_2 p_n\right)+\lambda\left(\sum_{i=1}^{n}p_i-1\right)$$

$$\frac{\partial L}{\partial p_i} = -\left(\log_2 p_i + p_i \times \frac{1}{\ln 2\, p_i}\right) + \lambda = 0, \quad i = 1,2,\ldots,n$$

$$\frac{\partial L}{\partial \lambda} = \sum_{i=1}^{n} p_i - 1 = 0 \qquad (*)$$

Solving the partial derivative above, we get the parameters for extreme value spot:

$$\lambda = \log_2 p_i + \frac{1}{\ln 2}$$

$p_i = 2^{\lambda - \frac{1}{\ln 2}}$ substituted into expression (*), we have

$$p_i = \frac{1}{n}$$

Thus, the maximum value is

$$I(P) = -\left(n \times \frac{1}{n}\log_2 \frac{1}{n}\right) = \log_2 n$$

(2) Lower bound

Let $g(p_i) = -p_i \log_2 p_i$,

so $g(p_i) \geq 0$ when $p_i \in [0,1]$ (calculated by 1st- and 2nd-order derivative)

Then

$$I(P) = -\left(p_1 \log_2 p_1 + p_2 \log_2 p_2 + \cdots + p_n \log_2 p_n\right) = \sum_{i=1}^{n} g(p_i) \geq 0$$

When anyone of $p_i$ is equal to 1 and others are equal to 0, we can have $I(P) = 0$, $\sum_{i=1}^{n} p_i - 1 = 0$. So, $I(P) = 0$ is the lower bound under the constraints.

Therefore, $p_i = \frac{1}{n}$ gives the max value of $I(P)$. $I(P) \in [0, \log_2 n]$.

3.2

(1) Upper bound

$$\text{Gini}(P) = 1 - \sum_{i=1}^{n} p_i^2$$

$$s.t. \sum_{i=1}^{n} p_i - 1 = 0$$

$$0 \le p_i \le 1$$

$$L(P,\beta) = 1 - \sum_{i=1}^{n} p_i^2 + \beta\left(\sum_{i=1}^{n} p_i - 1\right)$$

$$\frac{\partial L}{\partial p_i} = -2p_i + \beta = 0, \quad i = 1, 2, \dots, n$$

So,

$$p_i = \frac{\beta}{2}$$

$$\sum_{i=1}^{n} p_i = \frac{n\beta}{2} = 1$$

$$\sum_{i=1}^{n} p_i = \frac{n\beta}{2} = 1$$

$$p_i = \frac{1}{n}$$

Thus, the maximum value is $\text{Gini}(P) = 1 - n \times \frac{1}{n^2} = 1 - \frac{1}{n}$

(2) Lower bound

Let $\text{g}(p_i) = p_i^2$

$$\min \text{Gini}(P) = 1 - \sum_{i=1}^{n} g(p_i) \rightarrow \max \sum_{i=1}^{n} g(p_i)$$

Since $p_i^2 \leq p_i$ $(0 \leq p_i \leq 1)$

So, $\sum_{i=1}^{n} p_i^2 \leq \sum_{i=1}^{n} p_i$ $(0 \leq p_i \leq 1)$

$$\text{Gini}(P) = 1 - \sum_{i=1}^{n} p_i^2 \geq 1 - \sum_{i=1}^{n} p_i = 0$$

When arbitrary $p_i = 1$ and others are equal to 0, $\text{Gini}(P) = 1 - p_i = 0$.

Thus, $\text{Gini}(P) = 0$ is the lower bound.

Therefore, $p_i = \frac{1}{n}$ gives the max value of $Gini(P)$. $Gini(P) \in [0, 1 - \frac{1}{n}]$.

4.

$$\min_{w,b} \frac{1}{2} \| w \|^2$$

$$\text{s.t. } y_i \left( w^T x_i + b \right) \geq 1 \quad i = 1, 2, \ldots, m$$

First, we can change the constraints as:

$$y_i \left( w^T x_i + b \right) \geq 1 \quad i = 1, 2, \ldots, m$$

$$1 - y_i \left( w^T x_i + b \right) \leq 0 \quad i = 1, 2, \ldots, m$$

Then, we can construct the Lagrangian for our optimization problem:

$$L(w, b, \alpha) = \frac{1}{2} \| w \|^2 + \sum_{i=1}^{m} \alpha_i \left[ 1 - y_i \left( w^T x_i + b \right) \right]$$

Then, we need to get the derivatives of L respect to w and b to zero

$$\frac{\partial}{\partial w} L(w, b, a) = w - \sum_{i=1}^{m} \alpha_i y_i x_i = 0$$

$$so, w = \sum_{i=1}^{m} \alpha_i y_i x_i$$

$$\frac{\partial}{\partial b} L(w,b,a) = -\sum_{i=1}^{m} \alpha_i y_i = 0$$

$$so, \sum_{i=1}^{m} \alpha_i y_i = 0$$

Then, we need to plug them back into Lagrangian equation:

$$L(w,b,\alpha) = \frac{1}{2} \| w \|^2 + \sum_{i=1}^{m} \alpha_i \left[ 1 - y_i \left( w^T x_i + b \right) \right]$$

$$= \frac{1}{2} \sum_{i=1}^{m} \alpha_i y_i x_i \sum_{j=1}^{m} \alpha_j y_j x_j + \sum_{i=1}^{m} \alpha_i \left[ 1 - y_i \left( \sum_{j=1}^{m} \alpha_j y_j x_j^T x_i + b \right) \right]$$

$$= \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \left( \alpha_i \alpha_j y_i y_j x_i^T x_j \right) + \sum_{i=1}^{m} \alpha_i - \sum_{i=1}^{m} \alpha_i y_i \left( \sum_{j=1}^{m} \alpha_j y_j x_j^T x_i \right) - \sum_{i=1}^{m} \alpha_i y_i b$$

$$= \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \left( \alpha_i \alpha_j y_i y_j x_i^T x_j \right) + \sum_{i=1}^{m} \alpha_i - \sum_{i=1}^{m} \sum_{j=1}^{m} \left( \alpha_i \alpha_j y_i y_j x_i^T x_j \right) - b \sum_{i=1}^{m} \alpha_i y_i$$

$$= \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \left( \alpha_i \alpha_j y_i y_j x_i^T x_j \right)$$

Then, when we consider the dual problem:

$$\theta_D(\alpha) = \min_{w,b} L(w,b,\alpha)$$

$$\max_{\alpha, \alpha_i \geq 0} \theta_D(\alpha) = \max_{\alpha, \alpha_i \geq 0} \min_{w,b} L(w,b,\alpha)$$

So, we get the dual optimization problem:

$$\max_{\alpha} \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \left( \alpha_i \alpha_j y_i y_j x_i^T x_j \right)$$

$$\text{s.t. } a_i \geq 0 \quad i = 1, 2, \ldots, m$$

$$\sum_{i=1}^{m} \alpha_i y_i = 0$$

5.

When we get the derivatives of L respect to w and b to zero, we get

$$w = \sum_{i=1}^{m} \alpha_i y_i x_i$$

When we get the $\alpha^*$ just plug it in, we get

$$w^* = \sum_{i=1}^{m} \alpha_i^* y_i x_i$$

According to KKT condition, we get

$$\alpha_i^* g_i\left(w^*, b^*\right) = 0 \quad i = 1, 2, \ldots, m$$

When $\alpha_j^* > 0$, $g_j\left(w^*, b^*\right) = 1 - y_j\left(w^T x_j + b^*\right) = 0$

Put $w$ back into equation above:

$$1 - y_j\left(\left(\sum_{i=1}^{m} \alpha_i^* y_i x_i^T\right) x_j + b^*\right) = 0$$

$$\left(\left(\sum_{i=1}^{m} \alpha_i^* y_i x_i^T\right) x_j + b^*\right) = \frac{1}{y_j}$$

$$b^* = \frac{1}{y_j} - \sum_{i=1}^{m} \alpha_i^* y_i x_i^T x_j$$

Since $y_j \in \{1, -1\}$

Thus, $b^* = \frac{y_j^2}{y_j} - \sum_{i=1}^{m} \alpha_i^* y_i x_i^T x_j = y_j - \sum_{i=1}^{m} \alpha_i^* y_i x_i^T x_j$