SVM 1930026143

Q1.

The linear function used by a SVM for classification is:

$$y_i(w^T x_i + b) \geq 0$$

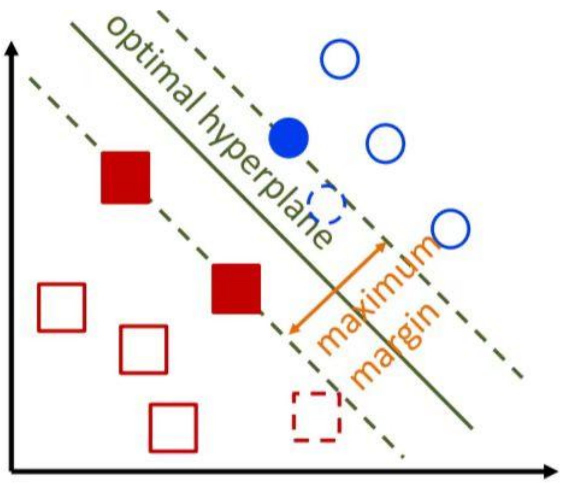
where:

$$y_i = \begin{cases} 1 & w^T x_i + b > 0 \\ 0 & w^T x_i + b > 0 \\ -1 & w^T x_i + b < 0 \end{cases}$$

An input x_i is positive if $w^T x_i + b > 0$, negative if $w^T x_i + b < 0$.

Q2.

If the training examples are linearly separable, it will have infinite decision boundaries can separate positive from negative data points because there is infinite $< w, b >$ pairs to express the plane, which have different norm of w ($||w||$). Eventually, we should select the points (support vectors) closest to the plane on the positive and negative sides of the hyperplane, so that the sum of the distances between these two points and the classification hyperplane (margin) is the largest. The distance of the Just like the graph:



The distance of the hyperplane and vectors d is $\frac{|w^T+b|}{||w||}$, so the hyperplane satisfies:

$$\max_{w,b}\left\{\min_{x\in D}\frac{2|w^T+b|}{||w||}\right\}$$

In this case, the generalization ability and robustness of the model are the strongest which means that it has strong tolerance and to local disturbance in training samples and adaptation to the other different dataset and tasks.

Q3.

3.1 For $P = (p_1, p_2, \ldots, p_n)$, $I(P) =- (p_1 \times \log p1 + p_2 \log p_2 + \cdots + p_n \log p_n)$.

$$\max I(P) = \max\ (-\sum_{k=1}^{n} p_k \log_2 p_k)$$

$$constrain: g(p_1, p_2, \ldots, p_n) = \sum_{k=1}^{n} p_k = 1$$

According to the Lagrange multiplier method:

$$L(p_1, p_2, \ldots, p_n, \alpha) = I(p_1, p_2, \ldots, p_n) + \alpha \cdot [g(p_1, p_2, \ldots, p_n) - 1] = 0$$

Take the partial derivative of each probability (p_k):

$$\frac{\partial F}{\partial p_k}\left(-\sum_{k=1}^{n} p_k \log_2 p_k + \alpha\left[\sum_{k=1}^{n} p_k - 1\right]\right) = 0$$

We can solve by the product rule:

$$\frac{\partial F}{\partial p_k}[-(p_k \log_2 p_k + \alpha p_k)] = 0$$

$$-\log_2 p_k - \frac{1}{In\ 2} + \alpha = 0$$

So we can get the solution:

$$p_k = 2^{\alpha - \frac{1}{In\ 2}}\ which\ is\ a\ constant$$

To conclude, all of probability (p_k) are identical with each other, which means that uniform distribution can get the maximum entropy.

3.2 For $P = (p_1, p_2, \ldots, p_n)$, $Gini(P) = 1 - (p_1^2 + p_2^2 + \cdots + p_n^2)$.

$$\max Gini(P) = \max\ (1 - \sum_{k=1}^{n} p_k^2)$$

$$constrain: g(p_1, p_2, \ldots, p_n) = \sum_{k=1}^{n} p_k = 1$$

According to the Lagrange multiplier method:

$$L(p_1, p_2, \ldots, p_n, \alpha) = Gini(p_1, p_2, \ldots, p_n) + \alpha \cdot [g(p_1, p_2, \ldots, p_n) - 1] = 0$$

Take the partial derivative of each probability (p_k):

$$\frac{\partial F}{\partial p_k}\left(-\sum_{k=1}^{n} p_k^2 + \alpha\left[\sum_{k=1}^{n} p_k - 1\right]\right) = 0$$

We can solve by the product rule:

$$\frac{\partial F}{\partial p_k}[-(p_k^2+\alpha p_k)]=0$$

$$-2\,p_k+\alpha=0$$

So we can get the solution:

$$p_k=\frac{\alpha}{2}\ \ which\ is\ a\ constant$$

To conclude, all of probability (p_k) are identical with each other, which means that uniform distribution can get the maximum Gini index value.

Q4.

Initially, we get the SVM optimization problem:

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

$$s.t.\ \ 1-y_i(w^Tx_i+b)\le 0,\qquad i=1,2,\dots,m$$

Which is a constrained optimization problem, then we can use the Lagrange multiplier method to transform it to the unconstrained form.

$$L(w,b,\alpha)=\frac{1}{2}||w||^2+\sum\nolimits_{i=1}^{N}\alpha_i\big(1-y_i(w^Tx_i+b)\big),\ \ \alpha_i\ge 0$$

So we get another expression to solve the original question:

$$\max_{\alpha}\frac{1}{2}||w||^2+\sum\nolimits_{i=1}^{N}\alpha_i\big(1-y_i(w^Tx_i+b)\big),\ \ \alpha_i\ge 0$$

When it satisfies the constrain, the result of $\max_{\alpha}L(w,b,\alpha)$ is just $\frac{1}{2}||w||^2$ or else it become to ∞ since $\alpha\ge 0$. The original problem can be converted to:

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

However, the solving process of the minimum-maximum problem is complicated and non-convex, the dual problem is used to help solve the problem (KKT condition must be satisfied to make the two problems equivalent):

$$\max_{\alpha}\min_{w,b}L(w,b,\alpha)=\max_{\alpha}\min_{w,b}\frac{1}{2}||w||^2+\sum\nolimits_{i=1}^{N}\alpha_i\big(1-y_i(w^Tx_i+b)\big)$$

Then find the partial derivative with respect to w and b respectively:

$$\frac{\partial L}{\partial w}=w-\sum\nolimits_{i=1}^{N}\alpha_iy_ix_i\ \ =0$$

$$\frac{\partial L}{\partial b} = \sum_{i=1}^{N} - \alpha_i y_i = 0$$

Thus, we can get: $w = \sum \alpha_i y_i x_i$, $\sum - \alpha_i y_i = 0$, substitute into the original:

$$\frac{1}{2}\left||w|\right|^2 + \sum_{i=1}^{N} \alpha_i\left(1 - y_i(w^T x_i + b)\right)$$

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

$$= \sum_{i=1}^{N} \alpha_i - \frac{1}{2}\sum_{i=1}^{N} \sum_{i=j}^{N} \alpha_i \alpha_j y_i y_j x_i^T x_j$$

Then we should make this part maximum:

$$\max_{\alpha} \sum_{i=1}^{N} \alpha_i - \frac{1}{2}\sum_{i=1}^{N} \sum_{i=j}^{N} \alpha_i \alpha_j y_i y_j x_i^T x_j$$

$$constrain\text{:} \sum_{i=1}^{N} - \alpha_i y_i = 0, \alpha_i \geq 0$$

To conclude, after a series of derivations, the dual form of SVM we get is the above.

Q5.

According to the KKT condition, we have:

$$\frac{\partial}{\partial W} L(W^*, b^*, \alpha^*) = W^* - \sum_{i=1}^{N} \alpha_i y_i x_i$$

$$\frac{\partial}{\partial b} L(W^*, b^*, \alpha^*) = W^* - \sum_{i=1}^{N} \alpha_i y_i x_i$$

$$\alpha^*\left(1 - y_i(w^T x_i + b)\right) = 0$$

$$1 - y_i(w^T x_i + b) \leq 0$$

$$\alpha_i \geq 0$$

So we can get $W^* = \sum_{i=1}^{N} \alpha_i^* y_i x_i$.

We must have a $\alpha_i^* \geq 0$ and it also satisfies:

$$y_j\left(w^{*T} x_j + b^*\right) = 1$$

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

And $y_j = \pm 1$ $(y_j = \frac{1}{y_j})$:

$$b_j = y_j - \sum_{i} \alpha_i^* y_i x_i^T x_j$$

To conclude,

- $w^* = \sum_{i=1}^{m} \alpha_i^* y_i \, \mathbf{x}_i$
- $b^* = y_j - \sum_{i=1}^{m} \alpha_i^* y_i \, \mathbf{x}_i{}^T \mathbf{x}_j$