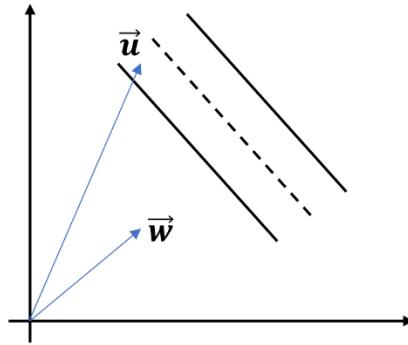


支持向量机的原理：



① 并不知道 \vec{u} 的长度，想要判断 \vec{u} 在街的哪一侧，如何判断？

如果 $\vec{u} \cdot \vec{w} \geq c$ （因为不知道街边到原点的距离）

那么可以说， \vec{u} 落在街的右侧。

把 $\vec{u} \cdot \vec{w} \geq c$ 改下形式：

$\vec{w} \cdot \vec{u} + b \geq 0$ （ b 是某个常数）

② 如果向量 \vec{u} 满足： $\vec{w} \cdot \vec{u} + b \geq 0$ ，我们称它为正例。

问题是，我们不知道 \vec{w} 是什么，不知道 b 应该怎么取值。（因为垂直于街道的 \vec{w} 有很多，它可以任意长）

为了求解 \vec{w} 和 b ，需要更多的约束条件。

增加约束：

对于正例 \vec{x}_+ ， $\vec{w} \cdot \vec{x}_+ + b \geq 1$

对于负例 \vec{x}_- ， $\vec{w} \cdot \vec{x}_- + b \leq -1$

为了数学上的便利性，增加变量 y_i 。对于正例， $y_i = 1$ ；对于负例， $y_i = -1$ 。

y_i 的取值，和正例还是负例有关。

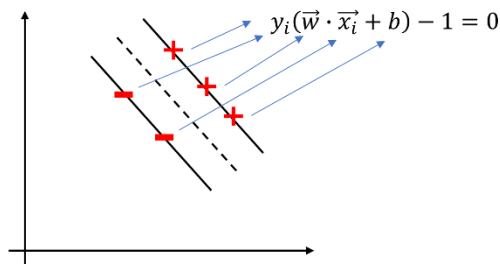
所以，可以把约束条件写为：

对于正例 \vec{x}_+ ： $y_i(\vec{w} \cdot \vec{x}_+ + b) \geq 1$

对于负例 \vec{x}_- ： $y_i(\vec{w} \cdot \vec{x}_- + b) \geq 1$

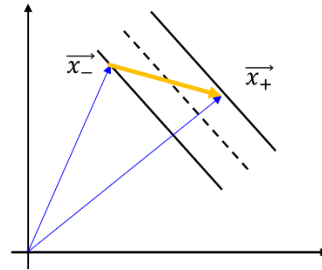
也就是说，对于所有例 \vec{x} ，都有 $y_i(\vec{w} \cdot \vec{x}_+ + b) - 1 \geq 0$ 。

继续增加约束：当例 \vec{x} 落在街边时， $y_i(\vec{w} \cdot \vec{x}_+ + b) - 1 = 0$ 。



问题的目的，是用街道把正负例分离，并且街道尽可能宽。
街道的宽度可以由落在街边上的例表示：

$$\text{Width} = (\vec{x}_+ - \vec{x}_-) \cdot \frac{\vec{w}}{\|\vec{w}\|}$$



对于正例，有： $\vec{w} \cdot \vec{x}_+ + b = 1$

对于负例，有： $\vec{w} \cdot \vec{x}_- + b = -1$

③ 所以， $\text{width} = \frac{1 - b + b + 1}{\|\vec{w}\|} = \frac{2}{\|\vec{w}\|}$

所以，问题的目的成了将 $\frac{2}{\|\vec{w}\|}$ 最大化，即将 $\|\vec{w}\|$ 最小化，即将 $\frac{1}{2} \|\vec{w}\|^2$ 最小化。（为了数学上的便利，选择二范数的平形式）怎么解呢？

思考还有什么约束条件？

之前约束了，在街边上， $y_i(\vec{w} \cdot \vec{x}_i + b) = 1$ 。所以，可以用拉格朗日乘子法将最值问题约束为：

$$\mathcal{L} = \frac{1}{2} \|\vec{w}\|^2 - \sum \alpha_i [y_i(\vec{w} \cdot \vec{x}_i + b) - 1]$$

其中， α_i 不能全为 0。

则：

$$\frac{\partial \mathcal{L}}{\partial \vec{w}} = \vec{w} - \sum \alpha_i y_i \vec{x}_i = 0$$

$$\frac{\partial \mathcal{L}}{\partial b} = \sum \alpha_i y_i = 0$$

④ 即：

$$\vec{w} = \sum \alpha_i y_i \vec{x}_i$$

$$\sum \alpha_i y_i = 0$$

（数学语言翻译 1： \vec{w} 是由样本集中位于街边上的 \vec{x}_i 和拉格朗日乘子集 α_i 决定的，即 \vec{w} 是样本集的线性组合。

数学语言翻译 2： α_i 不能全为 0，同时， α_i 要满足 $\sum \alpha_i y_i = 0$ ）

将

⑤ $\vec{w} = \sum \alpha_i y_i \vec{x}_i$ 代入 \mathcal{L} 中：

$$\begin{aligned}\mathcal{L} &= \frac{1}{2} \sum_{j=1}^N \sum_{i=1}^N \alpha_i \alpha_j y_i y_j \vec{x}_i \cdot \vec{x}_j - \sum_{i=1}^N \alpha_i \left[y_i \left(\sum_{j=1}^N \alpha_j y_j \vec{x}_j \cdot \vec{x}_i + b \right) - 1 \right] \\ &= \frac{1}{2} \sum_{j=1}^N \sum_{i=1}^N \alpha_i \alpha_j y_i y_j \vec{x}_i \cdot \vec{x}_j - \sum_{i=1}^N \alpha_i y_i \left(\sum_{j=1}^N \alpha_j y_j \vec{x}_j \cdot \vec{x}_i \right) - \sum_{i=1}^N \alpha_i y_i b + \sum_{i=1}^N \alpha_i \\ &= -\frac{1}{2} \sum_{j=1}^N \sum_{i=1}^N \alpha_i \alpha_j y_i y_j \vec{x}_i \cdot \vec{x}_j - 0 + \sum_{i=1}^N \alpha_i \\ &= -\frac{1}{2} \sum_{j=1}^N \sum_{i=1}^N \alpha_i \alpha_j y_i y_j \vec{x}_i \cdot \vec{x}_j + \sum_{i=1}^N \alpha_i\end{aligned}$$

观察发现， \mathcal{L} 的取值，和样本集之间的联系，仅仅是和样本集中的两个样本的内积有关。

⑥ 回到正负例的判断上：

$$\vec{w} \cdot \vec{x}_+ + b \geq 1$$

如果 $\sum \alpha_i y_i \vec{x}_i \cdot \vec{u} + b \geq 1$ ，则 \vec{u} 是正例。

⑦ 如果样本集是线性不可分的，则将其映射到另一个空间中，即通过坐标变换的方式使 \vec{x}_i 、 $\vec{x}_j \rightarrow \Phi(\vec{x}_i)$ 、 $\Phi(\vec{x}_j)$ 。映射后的 $\Phi(\vec{x}_i)$ 是线性可分的。

所以正负例的判断变为：

$$\sum \alpha_i y_i \Phi(\vec{x}_i) \cdot \Phi(\vec{u}) + b \geq 1$$

则 \vec{u} 是正例。

因为影响最终判断结果的是 $\Phi(\vec{x}_i) \cdot \Phi(\vec{u})$ ，所以未必要去找 Φ 映射，只要能找到 $K(\vec{x}_i, \vec{u}) = \Phi(\vec{x}_i) \cdot \Phi(\vec{u})$ 就好了。K 就称为 kernel function。

SVM 的计算过程总结：

- ① 先有了数据集，然后设了一个 b 。
- ② 求街道最宽的宽度，发现 $\vec{w} = \sum \alpha_i y_i \vec{x}_i$, $\sum \alpha_i y_i = 0$ 。 (α_i 不全为 0, α_i 的值还没确定，但是知道，存在这样的 α_i 使街道最宽，同时 α_i 满足： $\sum \alpha_i y_i = 0$ 。)
- ③ 正负例的判断标准变为：
若 \vec{u} 使： $\sum \alpha_i y_i \Phi(\vec{x}_i) \cdot \Phi(\vec{u}) + b \geq 1$ ，则 \vec{u} 为正例。
若 \vec{u} 使： $\sum \alpha_i y_i \Phi(\vec{x}_i) \cdot \Phi(\vec{u}) + b \leq -1$ ，则 \vec{u} 为负例。
并且，不用找 Φ ，找到 $K = \Phi \cdot \Phi$ 就行了。
需要找的是 α_i 。 α_i 需要满足的条件有：
 - (1) α_i 不全为 0；
 - (2) α_i 满足 $\sum \alpha_i y_i = 0$ ；
 - (3) α_i 的取值要满足样本集的分类结果。这样便可以寻找到适合的 α_i 。

如果计算 b 呢？

对于街边上的正例、反例：

$$y_i(\vec{w} \cdot \vec{x}_i + b) - 1 = 0$$

而 $y_i^2 = 1$ ，所以上式可以写为：

$$y_i(\vec{w} \cdot \vec{x}_i + b - y_i) = 0$$

所以：

$$\vec{w} \cdot \vec{x}_i + b - y_i = 0$$

$$b = y_i - \vec{w} \cdot \vec{x}_i = y_i - \sum_{j=1}^N \alpha_j y_j \vec{x}_j \cdot \vec{x}_i$$

(数学语言翻译： b 的值是由样本集和 α_i 所决定的。确定了 α_i 后， b 可以由样本集上，位于街边的任一例来生成得到。任一例对应着上式中的 \vec{x}_i ，样本集对应着上式中的 \vec{x}_j 。)

综上，SVM 只依赖于训练样本和样本内积，不依赖于高维映射函数，因为有核函数的存在。其他的全靠训练 α_i 来得到。然后可以使用 α_i 预测新样本 \vec{u} 的类别。预测方法如下：

$$\text{sign} \left(\sum \alpha_i y_i K(\Phi(\vec{x}_i) \cdot \Phi(\vec{u})) + (y_i - \sum_{j=1}^N \alpha_j y_j \vec{x}_j \cdot \vec{x}_{i-\text{random}}) \right) \rightarrow \begin{cases} < 0; \text{负例} \\ > 0; \text{正例} \end{cases}$$

(正例： $y_i(\vec{w} \cdot \vec{x}_i + b) \geq 1$ ；负例： $\vec{w} \cdot \vec{x}_i + b \leq -1$ 。所以由符号就能得出例的类别。)

所以，使用 SVM 完成分类问题，其实是根据样本集和 α_i 的约束条件，对 α_i 的寻找问题。