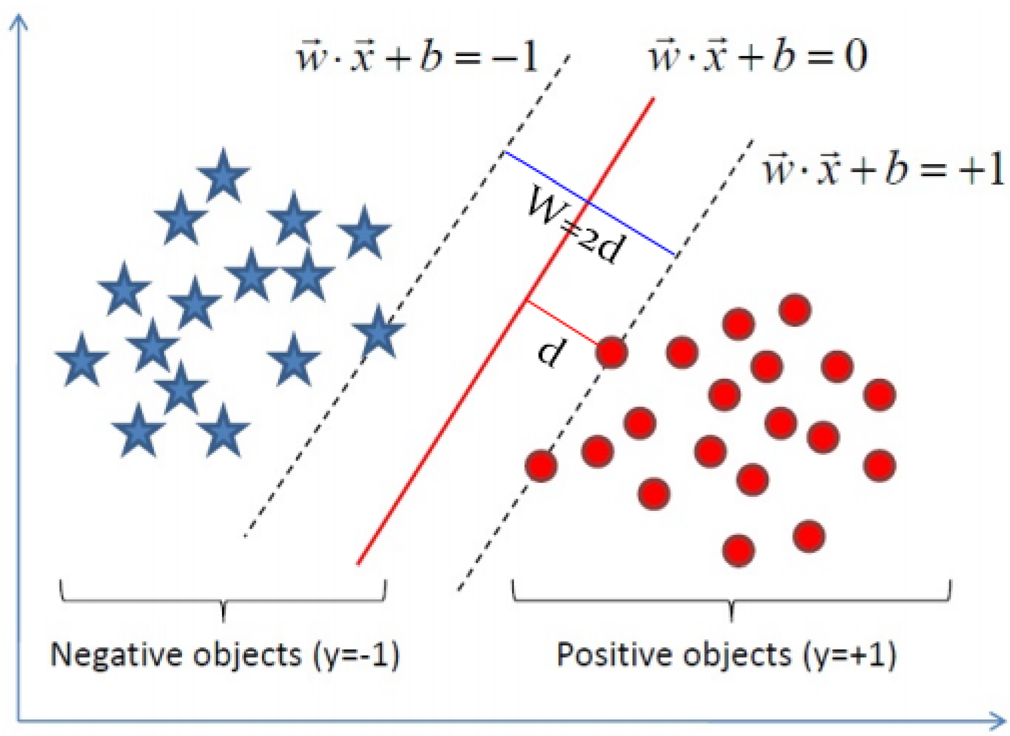
机器学习——支持向量机（SVM）复讲

之前做了支持向量机的笔记，过了这几天，觉得又有点忘了，重新拿起来复习一遍，然后作了一个更加精简的笔记，应该会更容易懂。

支持向量机在解决什么问题：

支持向量机和很多机器学习的算法一样，是针对分类的一种算法，最简单的理解就是线性二分类，如下图：



这副图其实就很概括性地理解了支持向量机的分类原理，什么样的分类间隔面才是合理并且最优的呢？显然是与两堆类别间隔最大的分隔面，说到底，SVM的原理就是一直在求、并且推导怎么得到这个最大间隔的分类面的参数，在这里，我们首先理解线性分类的参数，因为它是我们上手的基础。

间隔表示

为了最小化这个间隔，我们根据空间矢量的距离公式可以得到我们的目标函数（就是距离公式，不要说高中的距离公式都不会吧？可以看我上一篇具体有公式），然后由上图可以知道$(w^T x_i + b) = 1$，支持向量上的点满足图上表示的公式，代入其实就是最小化ω，没错，就是这么简单的目标，线性可分的支持向量机目标就是最小化这个ω。

当然我们还有约束条件，就是分类正确（如上图所示），应该满足：

$$s.t.\, y_i(w^T x_i + b) \geq 1$$

怎么会有这个公式？看图！分类正确就是满足这个公式，这个很好理解。

明显上面提供了两个条件，一个式子要最小化，一个约束条件，那么我们就遇到优化问题了。很好，拉格朗日函数来了。

拉格朗日函数优化：

之前看这个的时候，看到一篇写得很简单但很透彻的文章：

> https://www.jianshu.com/p/47986a0b1bf1

它讲拉格朗日式子为什么要这么构造讲得很好，其实懂了这个你根本不要看我上一篇列的式子，因为你根据上面的目标和约束条件，任何人都能构造出我们所要得到的函数，然后它也说到对偶性这个东西，我们就是根据这个转化这个式子，然后应该满足KKT条件，这个可以看我的上一篇具体讲述。

总之你只要懂得我们根据约束条件和目标函数列出拉格朗日函数，进行对偶、满足KKT、求导为零得到了：

$$\frac{\partial L}{\partial w}=0\rightarrow w=\sum_{1}^{n}\alpha_i y_i x_i$$
$$\frac{\partial L}{\partial b}=0\rightarrow \sum_{1}^{n}\alpha_i y_i=0$$

代入目标函数：

$$\max_{\alpha}\sum_{1}^{n}\alpha_i-\frac{1}{2}\sum_{i,j=1}^{n}\alpha_i\alpha_j y_i y_j x_i^T x_j$$
$$s.t.\alpha_i\geq 0, i=1,2,\ldots n$$
$$\sum_{1}^{n}\alpha_i y_i=0$$

没错，还是这么简单，就是求了导的含约束条件的目标函数，我们就是要进行alpha的求解。

SMO算法：

其实SMO算是比较复杂的一步了，因为前面都是简单的一步步顺承下来的，接下来是通过不断更新alpha来得到最后的结果。

接下来，因为我们通过两个alpha值来不断更新，所以把原来的目标函数进行拆分，然后化成最后的更新公式（这其中的都是简单的的换元推导）：

$$\alpha 2^{new}=\alpha_2^{old}+\frac{y_2(E_1-E_2)}{\eta}$$
$$E_i=f(x_i)-y_i$$
$$\eta=x_1^T x_1+x_2^T x_2-2x_1^T x_2$$

也就是说我们首先需要得到误差E，得到学习率eta，然后进行更新第二个参数，因为我们考虑了软判决（我们对分类给了一个可浮动的判决空间），所以添加了冗余量，使得我们需要对这个参数有一个上下限的修正，然后根据第二个参数更新第一个参数：

$$\alpha_1^{new}=\alpha_1^{old}+y_1 y_2(\alpha_2 old-\alpha_2^{new,clipped})$$

得到这两个参数后，我们就可以得到b：

$$b=\begin{cases} b_1 & 0<\alpha_1^{new}<C \\ b_2 & 0<\alpha_2^{new}<C \\ (b_1+b_2)/2 & otherwise \end{cases}$$

这就完成了！根据b算出w，ok！