

[登录](#) | [注册](#)

Rachel Zhang的专栏

因为理想，所以拼搏，关注计算机应用的点点滴滴……

[目录视图](#) [摘要视图](#) [RSS 订阅](#)

Rachel-Zhang

[赠书 | 异步2周年,技术图书免费选](#) [程序员8月书讯](#) [项目管理+代码托管+文档协作, 开发更流畅](#)

English (3)
Data Structure (75)
C/C (4)
php (2)
shell (3)
lua (0)
swig (1)
NLP (1)

阅读排行

SIFT特征提取分析 (385892)
机器学习——深度学习(C) (341320)
Stanford机器学习---第三 (163949)
初识压缩感知Compress (152063)
Matlab 线性拟合 & 非线性 (139699)
ROC曲线-阈值评价标准 (120168)
LibSVM 在matlab中的使 (112488)
GMM的EM算法实现 (96697)
压缩感知进阶——有关稀 (95328)
HOG特征-理解篇 (90974)

评论排行

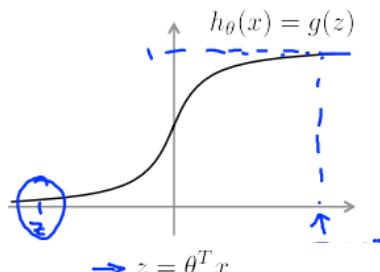
SIFT特征提取分析 (259)
初识压缩感知Compress (107)
机器学习——深度学习(C) (97)
一首诗, 致亲爱的csdn (90)
Stanford机器学习---第八 (90)
Stanford机器学习---第三 (89)
压缩感知进阶——有关稀 (84)
“压缩感知”之 “Hello Wor (80)
2012年度总结 (80)
GMM的EM算法实现 (80)

最新评论

双边滤波器的原理及实现
SWPU_Dreamer: 看了好多博主的文章, 写得质量都非常高,, 感谢, 感谢
K近邻分类算法实现 in Python
sinat_34022298:
@sinat_34022298: 可以参考这篇, 也是KNN的Python实现
http://blog....
ROC曲线-阈值评价标准
zbxzc: TPR定义有问题, 快改过来吧, 否则误导别人
车道检测问题探究(一)车道特征
jieshe1942: @xiangfeidicao: 请问你知道这个S0的值是如何设定的呢? 可否告知一下? 谢谢啦
车道检测问题探究(一)车道特征
jieshe1942: 请问这个S0的值是如何设定的呢?
opencv 中傅里叶变换 FFT
WenJi-Huang: 博主, 我看程序中用的是cvDFT, 这个不是快速傅里叶变化来的吧
2012年度总结
Bonaparte_cangtian: 现在小姑娘的技术水平让人很有压力阿, 不过《第三类爱情》《灵魂的台阶》这类都是层次很低的文学作品, 都能...

Alternative view of logistic regression

$$\rightarrow h_{\theta}(x) = \frac{1}{1 + e^{-\theta^T x}}$$



If $y = 1$, we want $h_{\theta}(x) \approx 1$, $\theta^T x \gg 0$
If $y = 0$, we want $h_{\theta}(x) \approx 0$, $\theta^T x \ll 0$

还是原先的假设, suppose我们只有两个类, $y=0$ 和 $y=1$ 。那么根据上图 $h(x)$ 的图形我们可以看出,

当 $y=1$ 时, 希望 $h(x) \approx 1$, 即 $z \gg 0$;

当 $y=0$ 时, 希望 $h(x) \approx 0$, 即 $z \ll 0$;

那么逻辑回归的cost function公式如下:

cost function我们之前已经讲过了, 这里不予赘述。现在呢, 我们来看看下面的两幅图, 这两幅图中灰色的curve是logistic regression的cost function分别取 $y=1$ 和 $y=0$ 的情况,

$y=1$ 时, 随着 $z \uparrow$, $h(x)$ 逐渐逼近1, cost逐渐减小。

$y=0$ 时, 随着 $z \downarrow$, $h(x)$ 逐渐逼近0, cost逐渐减小。

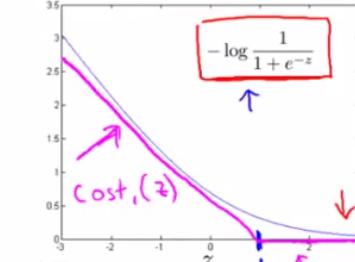
这正是图中灰色曲线所示的曲线。

ok, 现在我们来看看SVM中cost function的定义。请看下图中玫瑰色的曲线, 这就是我们希望得到的cost function曲线, 和logistic regression的cost function非常相近, 但是分为两部分, 下面呢, 我们将对这个cost function进行详细讲解。

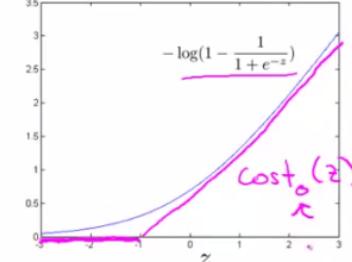
Alternative view of logistic regression (x, y)

$$\text{Cost of example: } -(y \log h_{\theta}(x) + (1-y) \log(1 - h_{\theta}(x))) \leftarrow \\ = -y \log \frac{1}{1 + e^{-\theta^T x}} - (1-y) \log(1 - \frac{1}{1 + e^{-\theta^T x}}) \leftarrow$$

If $y = 1$ (want $\theta^T x \gg 0$):
 $z = \theta^T x$



If $y = 0$ (want $\theta^T x \ll 0$):



logistic regression的cost function:

$$\min_{\theta} \frac{1}{m} \left[\sum_{i=1}^m y^{(i)} \left(-\log h_{\theta}(x^{(i)}) \right) + (1-y^{(i)}) \left(-\log(1 - h_{\theta}(x^{(i)})) \right) \right] + \frac{\lambda}{2m} \sum_{j=1}^n \theta_j^2$$

现在呢, 我们给出SVM的目标函数 (cost function) 定义:

$$\min_{\theta} C \sum_{i=1}^m \left[y^{(i)} \text{cost}_1(\theta^T x^{(i)}) + (1-y^{(i)}) \text{cost}_0(\theta^T x^{(i)}) \right] + \frac{1}{2} \sum_{j=1}^n \theta_j^2$$

该式中, cost_0 和 cost_1 分别对应 $y=0$ 和 $y=1$ 时的目标函数定义, 最后一项regularization项和logistic regression中的类似。感觉系数少了什么? 是的, 其实它们的最后一项本来是一样的, 但是可以通过线性变换化简得到SVM的归一化项。

怡心愿景
三人我师: 小学弟佩服的五体投地大神才女
SIFT特征提取分析
短腿小柯基: 楼主你好 我想请教一下有木有可能特征点自己手动选取???
Python数据可视化——散点图
Fiona_yuyu: 楼主, 这种方式(文章第二种的with label): ax.scatter(data,data,c=1...)

博客推荐
研究者July
CAD Feature Learning
Scikit guide
Face Datasets
UCI Dataset
MATLAB Statistics Demo

文章存档
2016年12月 (1)
2016年04月 (1)
2016年02月 (1)
2016年01月 (1)
2015年10月 (4)

展开

(二)、SVM — Large Margin Classifier

本节给出一个简单的结论——SVM是一个large margin classifier。什么是margin呢？下面我们做点论证将在下一节中给出。

在引入margin之前，我们回顾一下上一节中的SVM cost function curve，如下图所示分别是 $y=1$ 时和 $y=0$ 时的cost function curve。

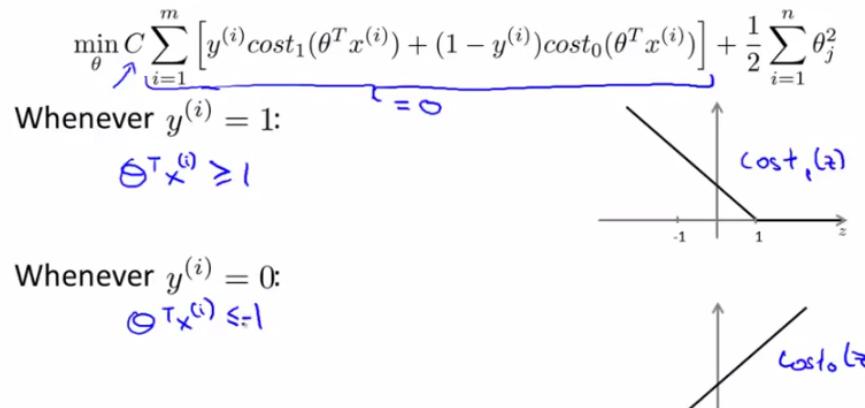
我们来看哈， C 很大，就要求 θ 中的那部分很小（令 θ 中的那部分表示为 W ），不如令其为0，这时来了：

※需求1：

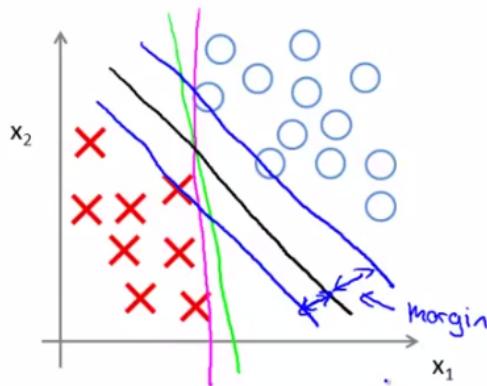
$y=1$ 时， W 只有前一项，令 $W=0$ ，就要求 $\text{Cost}_1(\theta^T x)=0$ ，由右图可知，这要求 $\theta^T x \geq 1$ ；

$y=0$ 时， W 只有后一项，令 $W=0$ ，就要求 $\text{Cost}_0(\theta^T x)=0$ ，由右图可知，这要求 $\theta^T x \leq -1$ ；

SVM Decision Boundary



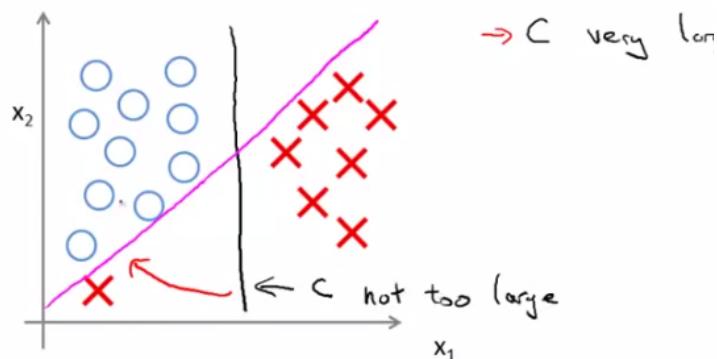
由以上说明可知，对 C 的取值应该在分类是否犯错和margin的大小上做一个平衡。那么 C 取较大的值会带来什么效果呢？就是我们开头说的结论——SVM是一个large margin classifier。那么什么是margin？在第三章中我们已经讲过了decision boundary，它是能够将所有数据点进行很好地分类的 $h(x)$ 边界。如下图所示，我们可以把绿色、粉色、蓝色或者黑色中的任意一条线当做decision boundary，但是哪一条最好呢？这里我们可以看出，绿色、粉色、蓝色这三类boundary离数据非常近，i.e. 我们再加进去几个数据点，很有可能这个boundary就能很好的进行分类了，而黑色的decision boundary距离两个类都相对较远，我们希望获得的就是这样的一个decision boundary。margin呢，就是将该boundary进行平移所得到的两条蓝线的距离，如图中所指。



Large margin classifier

相对比：

C小, decision boundary则呈现为黑线; 若C很大, 就呈现粉线;



这个结论大家可以记住, 也可以进行数学上的分析, 下一节中我们将从数学角度分析, 为什么SVM选用大value的C会形成一个large margin classifier。

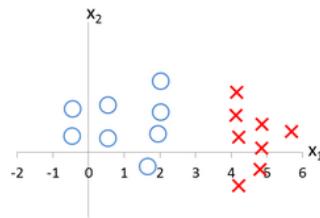
再给出一个数学上对geometry margin的说明:

$$\begin{aligned}
 x &= x_0 + y\gamma \frac{\mathbf{w}}{\|\mathbf{w}\|} \\
 \mathbf{w}^T \left(x - y\gamma \frac{\mathbf{w}}{\|\mathbf{w}\|} \right) + b &= 0 \\
 \gamma &= y \frac{\mathbf{w}^T x + b}{\|\mathbf{w}\|}
 \end{aligned}$$

任意一个点x到分类平面的距离 γ 的表示如上图所示, 其中y是{+1, -1}表示分类结果, x_0 是分类面上距x最短的点, 分类平面的方程为 $\mathbf{w}x+b=0$, 将 x_0 带入该方程就有上面的结果了。对于一个数据集x, margin就是这个数据及所有点的margin中离hyperplane最近的距离, SVM的目的就是找到最大margin的hyperplane。

练习:

Consider the training set to the right, where "x" denotes positive examples ($y = 1$) and "o" denotes negative examples ($y = 0$). Suppose you train an SVM (which will predict 1 when $\theta_0 + \theta_1 x_1 + \theta_2 x_2 \geq 0$). What values might the SVM give for θ_0 , θ_1 , and θ_2 ?



- $\theta_0 = 3, \theta_1 = 1, \theta_2 = 0$
- $\theta_0 = -3, \theta_1 = 1, \theta_2 = 0$
- $\theta_0 = 3, \theta_1 = 0, \theta_2 = 1$

Correct! $\times 3, \theta_1 = 0, \theta_2 = 1$

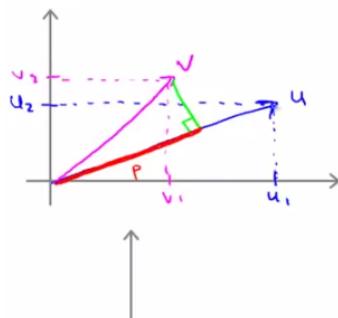
===== (三)、数学角度解析为什么SVM能形成 Large Margin Classifier (选看)

这一节主要为了证明上一节中的结论，为什么SVM是Large Margin Classification，能形成很好的decision boundary，如果仅仅处于应用角度考虑的朋友可以略过此节。

首先我们来看两个向量内积的表现形式。假设向量 u, v 均为二维向量，我们知道 u, v 的内积 $u^T v = u_1 v_1 + u_2 v_2$ 。表现在坐标上呢，就如下图左边所示：

首先将 v 投影至 u 向量，记其长度为 p （有正负，与 u 同向为正，反相为负，标量），则两向量的内积 $u^T v = \|u\| \cdot \|v\| \cdot \cos\theta = \|u\| \cdot p = u_1 v_1 + u_2 v_2$ 。

Vector Inner Product



$$\begin{aligned} u &= \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \rightarrow v = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \\ u^T v &=? \quad [u_1 \ u_2] \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \\ \|u\| &= \text{length of vector } u \\ &= \sqrt{u_1^2 + u_2^2} \in \mathbb{R} \\ p &= \text{length of projection of } v \text{ onto } u. \\ u^T v &= p \cdot \|u\| \leftarrow = v^T u \\ &= u_1 v_1 + u_2 v_2 \leftarrow \end{aligned}$$

这样一来，我们来看SVM的cost function:

$$\min_{\theta} C \sum_{i=1}^m \left[y^{(i)} \text{cost}_1(\theta^T x^{(i)}) + (1 - y^{(i)}) \text{cost}_0(\theta^T x^{(i)}) \right] + \frac{1}{2} \sum_{i=1}^n \theta_j^2$$

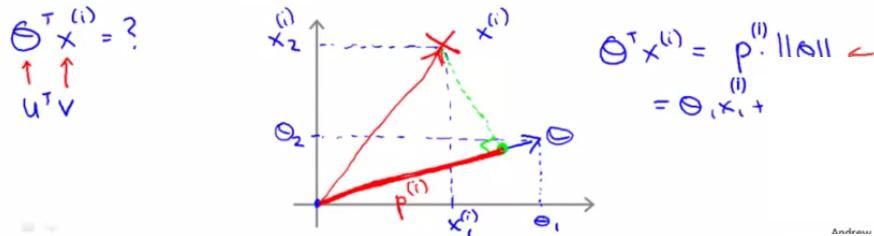
由于将 C 设的很大，cost function只剩下后面的那项。采取简化形式，意在说明问题即可，设 $\theta_0=0$ ，只剩下 θ_1 和 θ_2 ，

$$\text{则 cost function } J(\theta) = \frac{1}{2} \times \|\theta\|^2$$

SVM Decision Boundary

$$\min_{\theta} \frac{1}{2} \sum_{j=1}^n \theta_j^2 = \frac{1}{2} (\theta_1^2 + \theta_2^2) = \frac{1}{2} (\underbrace{\theta_1^2 + \theta_2^2}_{= \|\theta\|^2}) = \frac{1}{2} \|\theta\|^2$$

s.t. $\theta^T x^{(i)} \geq 1 \quad \text{if } y^{(i)} = 1$
 $\theta^T x^{(i)} \leq -1 \quad \text{if } y^{(i)} = 0$
Simplification: $\theta_0 = 0$, $n=2$



而根据上面的推导，有 $\theta^T x = p \cdot \|\theta\|$ ，其中 p 是 x 在 θ 上的投影，则

※需求2：

$y=1$ 时，W 只有前一项，令 $W=0$ ，就要求 $\text{Cost}_1(\theta^T x)=0$ ，由右图可知，这要求 $p \cdot \|\theta\| \geq 1$ ；

$y=0$ 时，W 只有后一项，令 $W=0$ ，就要求 $\text{Cost}_0(\theta^T x)=0$ ，由右图可知，这要求 $p \cdot \|\theta\| \leq -1$ ；

如下图所示：

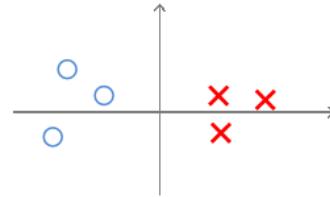
SVM Decision Boundary

$$\Rightarrow \min_{\theta} \frac{1}{2} \sum_{j=1}^n \theta_j^2 = \frac{1}{2} \|\theta\|^2 \leftarrow$$

s.t. $\underbrace{p^{(i)} \cdot \|\theta\| \geq 1}_{\text{if } y^{(i)} = 1}$
 $p^{(i)} \cdot \|\theta\| \leq -1 \quad \text{if } y^{(i)} = 0$
where $p^{(i)}$ is the projection of $x^{(i)}$ onto the vector θ .
Simplification: $\theta_0 = 0$

我们集中精力看为什么SVM的decision boundary有large margin（这里稍微有点儿复杂，好好看哈）：

对于一个给定数据集，依旧用X表示正样本，O表示负样本，绿色的线表示decision boundary，蓝色的线表示θ向量的方向，玫瑰色表示数据在θ上的投影。



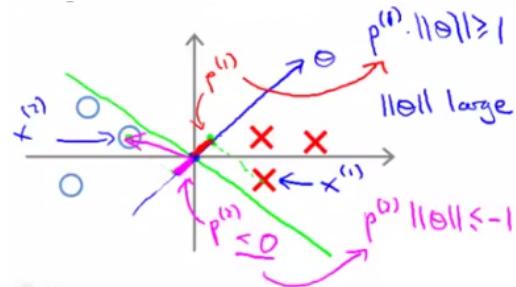
我们已知boundary的角度和θ向量呈的是90°角（自己画一下就知道了）。

先看这个图，对于这样一个decision boundary（没有large margin），θ与其呈90°角如图所示，这样我们可以画出数据集X和O在θ上的投影，如图所示，非常小；如果想满足[需求2]中说的

对正样本 $p \cdot \|\theta\| \geq 1$ ，

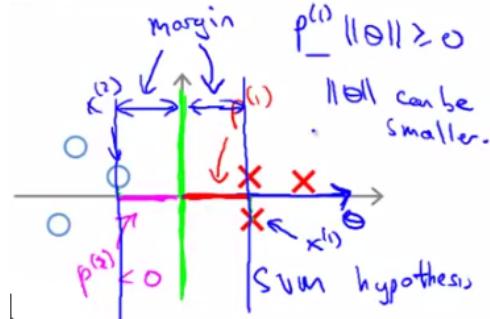
对负样本 $p \cdot \|\theta\| \leq -1$ ，

就需要令 $\|\theta\|$ 很大，这就和cost function的愿望 ($\min 1/2 \times \|\theta\|^2$) 相违背了，因此SVM的不出来这个图中所示的decision boundary结果。



那么再来看下面这个图，

它选取了上一节中我们定义的“比较好的”decision boundary，两边的margin都比较大。看一下它的投影，都比较大，这样就可以使 $\|\theta\|$ 相对较小，满足SVM的cost function。因此按照SVM的cost求解(optimization)得出的decision boundary一定是有large margin的。说明白了吧？！

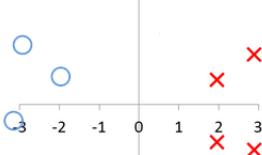


练习：

The SVM optimization problem we used is:

$$\min_{\theta} \frac{1}{2} \sum_{j=1}^n \theta_j^2$$

s.t. $\|\theta\| \cdot p^{(i)} \geq 1 \quad \text{if } y^{(i)} = 1$
 $\|\theta\| \cdot p^{(i)} \leq -1 \quad \text{if } y^{(i)} = -1$



where $p^{(i)}$ is the (signed - positive or negative) projection of $x^{(i)}$ onto θ . Consider the training set above. At the optimal value of θ , what is $\|\theta\|$?

1/4

1/2

1

[Continue](#)

Correct!

分析：由图中我们可以看出，decision boundary的最优解是 $y=x$ ，这时所有数据集中的数据到 θ 上的投影最小值为2，换言之，想满足

对正样本 $p \cdot \|\theta\| \geq 1$,

对负样本 $p \cdot \|\theta\| \leq -1$,

只需要

对正样本 $2 \cdot \|\theta\| \geq 1$,

对负样本 $(-2) \cdot \|\theta\| \leq -1$,

因此需要 $\|\theta\| \geq 1/2$ ，本着令cost function最小的原则，我们可知 $\|\theta\| = 1/2$.

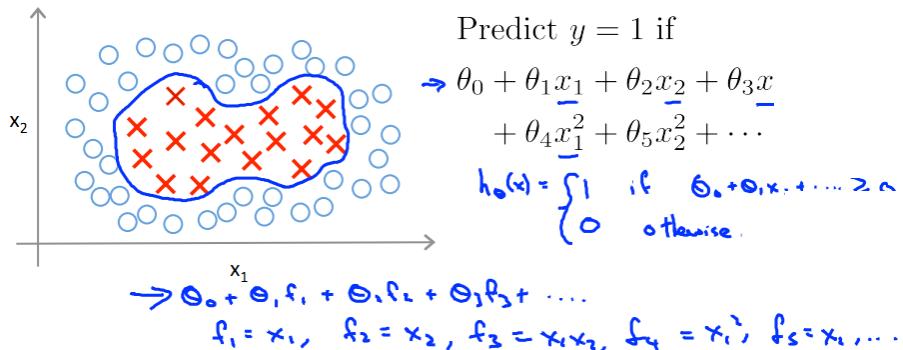
===== (四) 、SVM Kernel 1 — Gaussian Kernel

对于一个非线性Decision boundary，我们之前利用多项式拟合的方法进行预测：

- f_1, f_2, \dots, f_n 为提取出来的 features。
- 定义预测方程 $h_\theta(x)$ 为多项式的 sigmoid 函数值: $h_\theta(x) = g(\theta_0 f_0 + \theta_1 f_1 + \dots + \theta_n f_n)$, 其中 f_n 为 x 的幂次项组合 (如下图)
- 当 $\theta_0 f_0 + \theta_1 f_1 + \dots + \theta_n f_n \geq 0$ 时 $h_\theta(x) = 1$; else $h_\theta(x) = 0$;

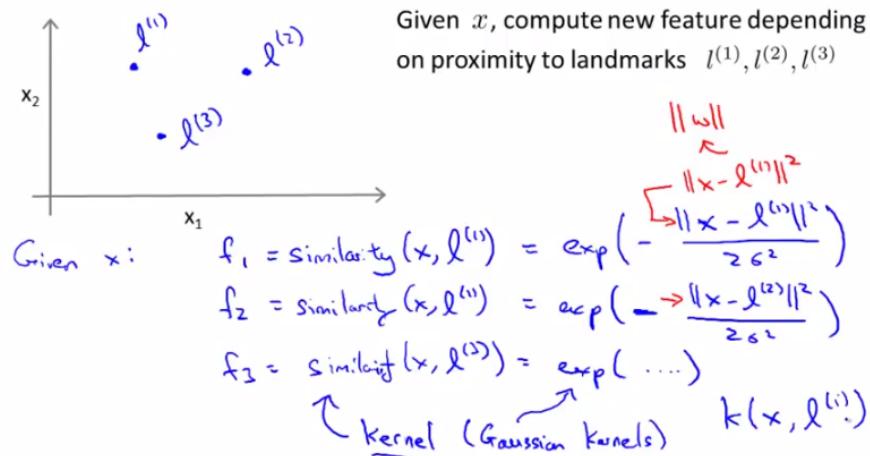
那么, 除了将 f_n 定义为 x 的幂次项组合, 还有没有其他方法表示 f 呢? 本节就引入了 Kernel, 核的概念。即用核函数表示 f 。

Non-linear Decision Boundary



对于上图的非线性拟合, 我们通过计算输入原始向量与 landmark 之间的相似度来计算核值 f :

Kernel



发现相似度计算公式很像正态分布 (高斯分布) 对不对? 是的! 这就是高斯核函数。由下图可以看出,

x 和 l 越相似, f 越接近于 1;

x 与 l 相差越远, f 越接近于 0;

Kernels and Similarity

$$f_1 = \text{similarity}(x, l^{(1)}) = \exp\left(-\frac{\|x - l^{(1)}\|^2}{2\sigma^2}\right) = \exp\left(-\frac{\sum_{j=1}^n (x_j - l_j^{(1)})^2}{2\sigma^2}\right)$$

If $x \approx l^{(1)}$:

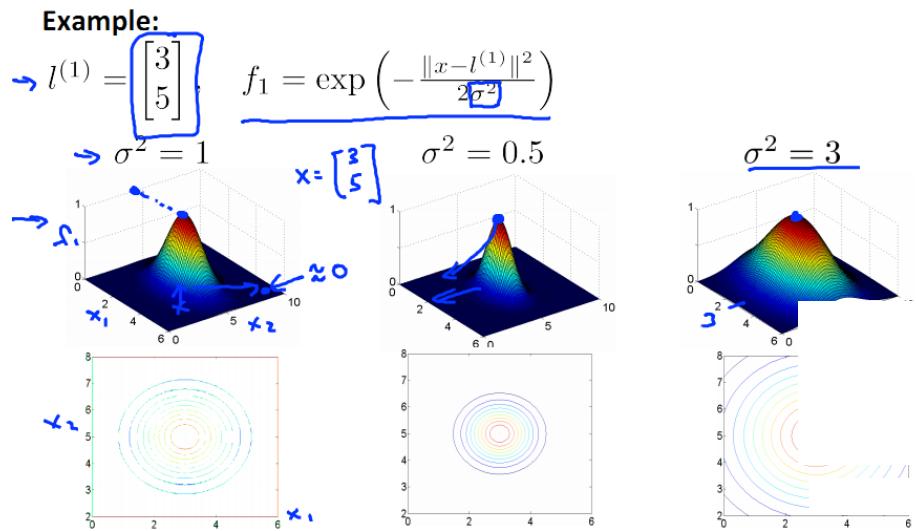
$$f_1 \underset{\uparrow}{\approx} \exp\left(-\frac{\underset{\downarrow}{0}}{2\sigma^2}\right) \approx 1$$

If x if far from $l^{(1)}$:

$$f_1 = \exp\left(-\frac{(\text{large number})^2}{2\sigma^2}\right) \approx 0.$$

下图中的横纵坐标为 x 的两个维度值, 高为 f (new feature)。制高点为 $x=l$ 的情况, 此时 $f=1$ 。

随着 x 与 l 的远离, f 逐渐下降, 趋近于 0。

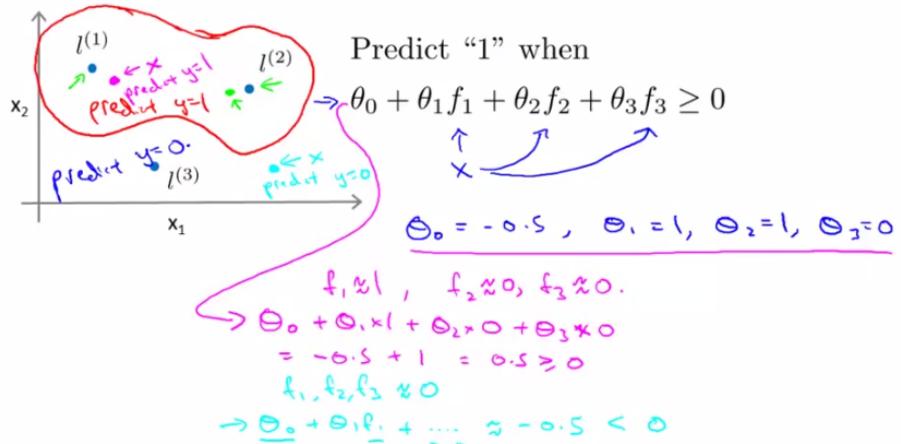


下面我们来看SVM核分类预测的结果：

引入核函数后，代数上的区别在于f变了，原来f是 $x_1/x_1^2/\dots$ ，即 x_i 幂次项乘积

引入核函数后，几何上来说可以更直观的表示是否应该归为该类了（如下图）

- 比如我们想将坐标上的所有数据点分为两类（如下图中）红色圈内希望预测为 $y=1$ ；圈外希望预测为 $y=0$ 。通过训练数据集呢，我们得到了一组 θ 值 $(\theta_0, \theta_1, \theta_2, \theta_3) = (-0.5, 1, 1, 0)$ 以及三个点 (L_1, L_2, L_3) ，（具体怎么训练而成的大家先不要过分纠结，后面会讲）
- 对于每个test数据集中的点，我们首先计算它到 (L_1, L_2, L_3) 各自的相似度，也就是核函数的值 (f_1, f_2, f_3) ，然后带入多项式 $\theta_0 f_0 + \theta_1 f_1 + \dots + \theta_n f_n$ 计算，当它 ≥ 0 时，预测结果为类内点（正样本， $y=1$ ），else预测为负样本， $y=0$



=====

(五)、SVM 中 Gaussian Kernel 的使用

§ 5.1. landmark的选取和参数向量θ的求解

上一节中我们遗留了两个问题，一个是一些L点的选取，一个是向量 θ 计算。这一节我们就来讲讲这两个问题。

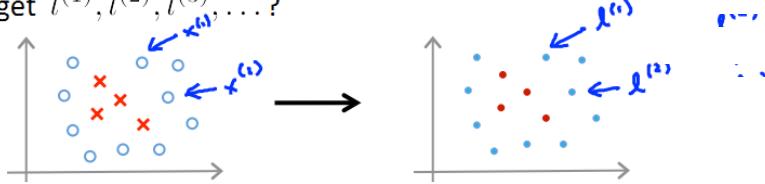
首先来看L的选取。上一节中提到Gaussian kernel f_i 的计算：

Given x :

$$\begin{aligned} f_i &= \text{similarity}(x, l^{(i)}) \\ &= \exp\left(-\frac{\|x - l^{(i)}\|^2}{2\sigma^2}\right) \end{aligned}$$

这里呢，我们选择m个训练数据，并取这m个训练数据为m个landmark (L) 点（不考虑样本x）
如下图所示：

Predict $y = 1$ if $\theta_0 + \theta_1 f_1 + \theta_2 f_2 + \theta_3 f_3 \geq 0$ ←
Where to get $f_1^{(1)}, f_2^{(2)}, f_3^{(3)}, \dots$?



SVM with Kernels

- Given $(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \dots, (x^{(m)}, y^{(m)})$,
- choose $l^{(1)} = x^{(1)}, l^{(2)} = x^{(2)}, \dots, l^{(m)} = x^{(m)}$.

Given example x :

- $f_1 = \text{similarity}(x, l^{(1)})$
- $f_2 = \text{similarity}(x, l^{(2)})$
- ...

$$f = \begin{bmatrix} f_0 \\ f_1 \\ f_2 \\ \vdots \\ f_m \end{bmatrix} \quad f_0 = 1$$

For training example $(x^{(i)}, y^{(i)})$:

$$\begin{aligned} x^{(i)} \rightarrow \boxed{f_1^{(i)} = \sin(x^{(i)}, l^{(1)})} \\ f_2^{(i)} = \sin(x^{(i)}, l^{(2)}) \\ \vdots \\ f_i^{(i)} = \sin(x^{(i)}, l^{(i)}) = \exp\left(-\frac{\|x^{(i)} - l^{(i)}\|^2}{2\sigma^2}\right) = 1 \\ f_{m+1}^{(i)} = \sin(x^{(i)}, l^{(m)}) \end{aligned}$$

$$\begin{aligned} x^{(i)} \in \mathbb{R}^{n+1} \quad (\text{or } \mathbb{R}^n) \\ f^{(i)} = \begin{bmatrix} f_0^{(i)} \\ f_1^{(i)} \\ f_2^{(i)} \\ \vdots \\ f_m^{(i)} \end{bmatrix} \\ f_0^{(i)} = 1 \end{aligned}$$

Andrew Ng

PS：那么在这m个训练数据中，每一个训练数据 $x(i)$ 所得的特征向量（核函数） f 中，总有一维向量的值为1（因为这里 $x(i)=l(i)$ ）

于是，每个特征向量 f 有 $m+1$ 维（ m 维训练数据 $[f_1, f_2, \dots, f_m]$ 附加一维 $f_0=1$ ）

在SVM的训练中，将Gaussian Kernel带入cost function，通过最小化该函数就可得到参数 θ ，并根据该参数 θ 进行预测：

```
若  $\theta^T f > 0$ , predict y=1;
else predict y=0;
```

如下图所示，这里与之前讲过的cost function的区别在于用kernel f代替了x。

SVM with Kernels

Hypothesis: Given x , compute features $f \in \mathbb{R}^{m+1}$

Predict “y=1” if $\theta^T f \geq 0$

Training:

$$\min_{\theta} C \sum_{i=1}^m y^{(i)} \text{cost}_1(\theta^T f^{(i)}) + (1 - y^{(i)}) \text{cost}_0(\theta^T f^{(i)}) + \frac{1}{2} \sum_{j=1}^n \theta_j^2$$

§ 5.2. landmark的选取和参数向量 θ 的求解

好了，至此Landmark点和 θ 的求取都解决了，还有一个问题，就是cost function中两个参数的确定： C 和 σ^2 。

对于C, 由于 $C=1/\lambda$, 所以

C 大, λ 小, overfit, 产生low bias, high variance

C 小, λ 大, underfit, 产生high bias, low variance

详细原因请参考第六章中关于bias和variance的讲解。

对于方差 σ^2 , 和正态分布中的定义一样,

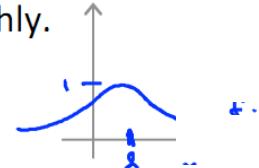
σ^2 大, $x-f$ 图像较为扁平;

σ^2 小, $x-f$ 图像较为窄尖;

Large σ^2 : Features f_i vary more smoothly.

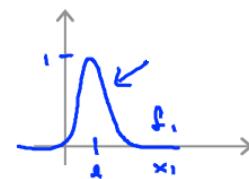
→ Higher bias, lower variance.

$$\exp\left(-\frac{\|x-\mu\|^2}{2\sigma^2}\right)$$



Small σ^2 : Features f_i vary less smoothly.

Lower bias, higher variance.



关于C和 σ^2 的选取, 我们来做个练习:

Suppose you train an SVM and find it overfits your training data. Which of these would be a reasonable next step? Check all that apply.

Increase C

Decrease C

Increase σ^2

Decrease σ^2

Continue

Correct!

解析, 过拟合说明应该适当加强cost function中的正则项所起的作用, 因此应增大 λ , 即减小C; 同时, 过拟合是的只有一小部分范围内的x享有较大f, 或者说x的覆盖面太窄了, 所以应当增大 σ^2 。

=====

(六)、SVM 的使用与选择

本节中主要介绍SVM在matlab中用libsvm中的应用, 给大家一个用SVM进行实践的平台。

前面几节中我们已知用SVM进行机器学习的过程就是一个optimize参数 θ 的过程, 这里呢, 我们首先介绍一个 Chih-Chung Chang 和 Chih-Jen Lin 做的 matlab/C/Ruby/[Python/Java](#)...中通用的机器学习tool, libsvm,

其基本讲解和测试我以前讲过（[在这里](#)），算是入门篇，并不详细，这里呢，我们将结合本章课程进一步学习，并用matlab实现。

首先大家来看看，想要进行SVM学习，有哪两类：

一种是No kernel (linear kernel) , $h_0(x)=g(\theta_0x_0+\theta_1x_1+\dots+\theta_nx_n)$, predict $y=1$ if $\theta^T x \geq 0$;

另一种是使用kernel f (比如Gaussian Kernel) , $h_0(x)=g(\theta_0f_0+\theta_1f_1+\dots+\theta_nf_n)$, 这里需要选择方差参数 σ^2 如下图所示：

Use SVM software package (e.g. liblinear, libsrm, ...) to solve for parameters θ .

Need to specify:

→ Choice of parameter C.

Choice of kernel (similarity function):

E.g. No kernel ("linear kernel")

Predict "y = 1" if $\underline{\theta}^T x \geq 0$

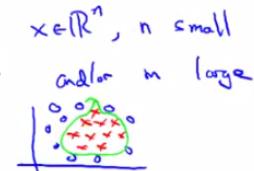
$$\theta_0 + \theta_1x_1 + \dots + \theta_nx_n \geq 0 \quad x \in \mathbb{R}^n$$

$\rightarrow n$ large, m small

Gaussian kernel:

$$f_i = \exp\left(-\frac{\|x - l^{(i)}\|^2}{2\sigma^2}\right), \text{ where } l^{(i)} = x^{(i)}.$$

Need to choose σ^2



需要注意的是，不管用那种方法，都需要在ML之前进行Normalization归一化！

当然，除了Gaussian kernel，我们还有很多其他的kernel可以用，比如polynomial kernel等，如下图所示，但andrew表示他本人不会经常去用（或者几乎不用）以下"more esoteric"中的核，一个原因是其他的核不一定起作用。我们讲一下polynomial kernel:

polynomial 核形如 $K(x, l) = (x^T l + c)^d$, 也用来表示两个object的相似度

Other choices of kernel

Note: Not all similarity functions $\text{similarity}(x, l)$ make valid kernels.

→ (Need to satisfy technical condition called "Mercer's Theorem" to make sure SVM packages' optimizations run correctly, and do not diverge).

Many off-the-shelf kernels available:

- Polynomial kernel: $k(x, l) = (x^T l + \text{constant})^{\text{degree}}$

$$k(x, l) = (x^T l)^2, (x^T l)^3, (x^T l + 1)^3, (x^T l + 5)^4$$

- More esoteric: String kernel, chi-square kernel, histogram intersection kernel, ...

$$\text{sim}(x, l)$$

首先给大家引入一个[数据集](#)，在该数据集中，我们可以进行初步的libsrm训练和预测，如[这篇文章](#)中所说，这个也是最基本的no kernel(linear kernel)。

然后呢，给大家一个reference，这是libsrm中train的基本的语法：

```
[cpp]
01. Usage: model = svmtrain(training_label_vector, training_instance_matrix, 'libsvm_option';
02. libsvm_options:
03. -s svm_type : set type of SVM (default 0)
04.   0 -- C-SVC
05.   1 -- nu-SVC
06.   2 -- one-class SVM
07.   3 -- epsilon-SVR
08.   4 -- nu-SVR
09. -t kernel_type : set type of kernel function (default 2)
```

```

10.    0 -- linear: u'*v
11.    1 -- polynomial: (gamma*u'*v + coef0)^degree
12.    2 -- radial basis function: exp(-gamma*|u-v|^2)
13.    3 -- sigmoid: tanh(gamma*u'*v + coef0)
14.    4 -- precomputed kernel (kernel values in training_instance_matrix)
15. -d degree : set degree in kernel function (default 3)
16. -g gamma : set gamma in kernel function (default 1/num_features)
17. -r coef0 : set coef0 in kernel function (default 0)
18. -c cost : set the parameter C of C-SVC, epsilon-SVR, and nu-SVR (default 1)
19. -n nu : set the parameter nu of nu-SVC, one-class SVM, and nu-SVR (default 0.5)
20. -p epsilon : set the epsilon in loss function of epsilon-SVR (default 0.1)
21. -m cachesize : set cache memory size in MB (default 100)
22. -e epsilon : set tolerance of termination criterion (default 0.001)
23. -h shrinking : whether to use the shrinking heuristics, 0 or 1 (default 1)
24. -b probability_estimates : whether to train a SVC or SVR model for probabil
25. -wi weight : set the parameter C of class i to weight*C, for C-SVC (default
26. -v n : n-fold cross validation mode
27. -q : quiet mode (no outputs)

```

下面给大家一个例子：

```

[cpp]

01. function [ output_args ] = Nonlinear_SVM( input_args )
02. %NONLINEAR_SVM Summary of this function goes here
03. % Detailed explanation goes here
04.
05. %generate data1
06. r=sqrt(rand(100,1));%generate 100 random radius
07. t=2*pi*rand(100,1);%generate 100 random angles, in range [0,2*pi]
08. data1=[r.*cos(t),r.*sin(t)];%points
09.
10. %generate data2
11. r2=sqrt(3*rand(100,1)+1);%generate 100 random radius
12. t2=2*pi*rand(100,1);%generate 100 random angles, in range [0,2*pi]
13. data2=[r2.*cos(t2),r2.*sin(t2)];%points
14.
15. %plot datas
16. plot(data1(:,1),data1(:,2),'r.')
17. hold on
18. plot(data2(:,1),data2(:,2),'b.')
19. ezpolar(@(x)1);%在极坐标下画p=1, θ∈[0, 2π]的图像, 即x^2+y^2=1
20. ezpolar(@(x)2);
21. axis equal %make x and y axis with equal scalar
22. hold off
23.
24. %build a vector for classification
25. data=[data1;data2]; %merge the two dataset into one
26. datalabel=ones(200,1); %label for the data
27. datalabel(1:100)=-1;
28.
29. %train with Non-linear SVM classifier use Gaussian Kernel
30.
31. model=svmtrain(datalabel,data,'-c 100 -g 4');
32.
33. end

```

该例中我们分别生成了100个正样本和100个负样本，如下图所示，因为kernel type default=2（即Gaussian kernel），通过`svmtrain(datalabel, data, '-c 100 -g 4')`我们设置了第五节中奖的参数——C (c) 和 $2\sigma^2$ (g) 分别为100和4。

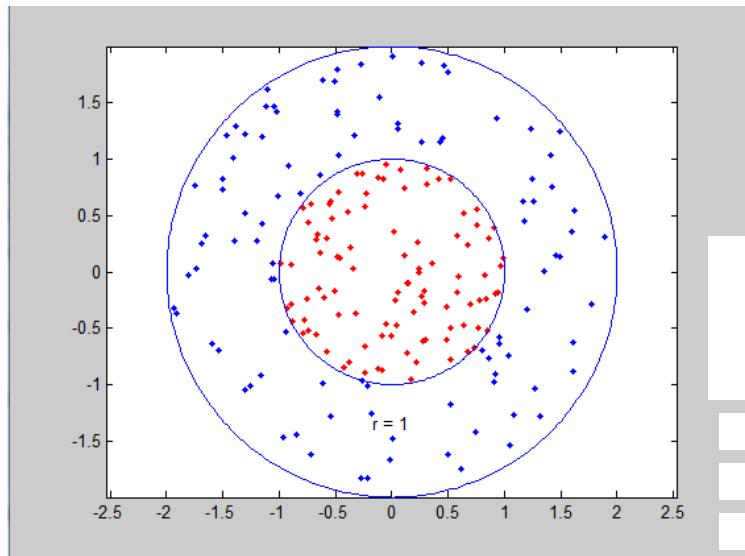
运行结果：

```

[cpp]

01. >> Nonlinear_SVM
02. *
03. optimization finished, #iter = 149
04. nu = 0.015538
05. obj = -155.369263, rho = 0.634344
06. nSV = 33, nBSV = 0
07. Total nSV = 33

```



最后，我们比较一下logistic regression和SVM：

用n表示feature个数，m表示training example个数。

①当 $n \geq m$ ，如 $n=10000$, $m=10 \sim 1000$ 时，建议用logistic regression, 或者linear kernel的SVM

②如果n小, m不大不小, 如 $n=1 \sim 1000$, $m=10 \sim 10000$, 建议用Gaussian Kernel的SVM

③如果n很小, m很大, 如 $n=1 \sim 1000$, $m>50000$, 建议增加更多的feature并使用logistic regression, 或者linear kernel的SVM

原因，①模型简单即可解决，③如果还用Gaussian kernel会导致很慢，所以还选择logistic regression或者linear kernel

神经网络可以解决以上任何问题，但是速度是一个很大的问题。

详见下图：

Logistic regression vs. SVMs

$n = \text{number of features } (x \in \mathbb{R}^{n+1})$, $m = \text{number of training examples}$

- If n is large (relative to m): (e.g. $n \geq m$, $n=10,000$, $m=10 \sim 1000$)
- Use logistic regression, or SVM without a kernel ("linear kernel")
- If n is small, m is intermediate: ($n=1 \sim 1000$, $m=10 \sim 10,000$)
- Use SVM with Gaussian kernel
- If n is small, m is large: ($n=1 \sim 1000$, $m=50,000+$)
- Create/add more features, then use logistic regression or SVM without a kernel
- Neural network likely to work well for most of these settings, but may be slower to train.



test:

Suppose you are trying to decide among a few different choices of kernel and are also choosing parameters such as C , σ^2 , etc. How should you make the choice?

- Choose whatever performs best on the training data.
- Choose whatever performs best on the cross-validation data.
- Choose whatever performs best on the test data.
- Choose whatever gives the largest SVM margin.

Correct! 

我们可以把所有数据分为testset和training set两部分进行训练，example：

```
[cpp]
01. load heart_scale
02. [N D] = size(heart_scale_inst);
03.
04. % Determine the train and test index, select top 200 as training data
05. % else as test data
06. trainIndex = zeros(N,1); trainIndex(1:200) = 1;
07. testIndex = zeros(N,1); testIndex(201:N) = 1;
08. trainData = heart_scale_inst(trainIndex==1,:);
09. trainLabel = heart_scale_label(trainIndex==1,:);
10. testData = heart_scale_inst(testIndex==1,:);
11. testLabel = heart_scale_label(testIndex==1,:);
12.
13. % Train the SVM
14. model = svmtrain(trainLabel, trainData, '-c 1 -g 0.07 -b 1');
15. % Use the SVM model to classify the data
16. [predict_label, accuracy, prob_values] = svmpredict(testLabel, testData, model, '-b 1');
```

运行结果：

```
[cpp]
01. optimization finished, #iter = 87
02. nu = 0.426369
03. obj = -56.026822, rho = -0.051128
04. nSV = 77, nBSV = 62
05. Total nSV = 77
06. *
07. optimization finished, #iter = 99
08. nu = 0.486493
09. obj = -64.811759, rho = 0.328505
10. nSV = 87, nBSV = 68
11. Total nSV = 87
12. *
13. optimization finished, #iter = 101
14. nu = 0.490332
15. obj = -64.930603, rho = 0.424679
16. nSV = 87, nBSV = 67
17. Total nSV = 87
18. *
19. optimization finished, #iter = 121
20. nu = 0.483649
21. obj = -64.046644, rho = 0.423762
22. nSV = 87, nBSV = 65
23. Total nSV = 87
24. *
25. optimization finished, #iter = 93
26. nu = 0.470980
27. obj = -63.270339, rho = 0.458209
28. nSV = 83, nBSV = 67
29. Total nSV = 83
```

```
30.  *
31. optimization finished, #iter = 137
32. nnu = 0.457422
33. obj = -76.730867, rho = 0.435233
34. nSV = 104, nBSV = 81
35. Total nSV = 104
36. Accuracy = 81.4286% (57/70) (classification)
37. >>
```

这里只是一部分我做过的实验，希望有朋友能够有更完善的程序或者更好的资料推荐~谢谢~

小结

本章讲述了Support Vector Machine的基本原理、SVM与linear regression、logistic regression、神经网络的关系和matlab中通过Libsvm库对数据进行训练，希望对大家有所帮助。

关于Machine Learning更多的学习资料将继续更新，敬请关注本博客和新浪微博[Sophia_q...@sina.com](#)。

Reference:

- 1.[How to build a custom Kernel function and use it with Libsvm in C?](#)
- 2.[Libsvm在matlab中的使用](#)
- 3.[SVM parameter tuning and number of SVs \(Matlab libsvm\)](#)
- 4.[Libsvm for matlab_Kittipat](#)

顶 踩

37

2

上一篇 [Python进阶（一）——安装Python、程序执行、Python模块和IDLE调试](#)

下一篇 [Python进阶（二）——Python对象类型](#)

相关文章推荐

- 机器学习实验(九):基于高斯分布和OneClassSV...
- 【直播】70天软考冲刺计划--任铄
- one class SVM
- 【直播】打通Linux脉络 进程、线程、调度--宋宝华
- One-class SVM
- 【直播】机器学习之凸优化--马博士
- One Class SVM, SVDD (Support Vector Domain ...
- 【套餐】MATLAB基础+MATLAB数据分析与统计--...
- two-class SVM & one-class SVM & exemplar SV...
- 【课程】3小时掌握Docker最佳实战--徐西宁
- OpenCV学习笔记(二十六)——小试SVM算法ml
- 【课程】深度学习基础与TensorFlow实践--AI100
- LibSVM学习详细说明
- libSVM简介及核函数模型选择
- SVM理解
- 手把手教你实现SVM算法(一)

[查看评论](#)

45楼 No_defeat 2017-03-20 16:13发表



log函数的底数是10?

44楼 stellar1 2017-02-17 09:46发表



引用"qq_20371377"的评论:

关于:

我有个问题:

你说"由于将C设的很大, cost function只剩下后面的那项。采取简化形...

不是不优化前项,而是正因为前项太重要了,所以干脆把前项直接写成限制条件(就是s.t.的部分)。因为当点,就比如C是正无穷吧),此时只要有哪怕一个点的cost的值不是0(即这个点没有落在正确分类那一侧的margin就是正无穷。此时无论怎么优化后项都完全没有意义了。要避免这种情况,就必须保证前项中所有的cost都是不是“优化目标”而是“限制条件”了,所以才把前项直接写成了s.t.,只剩下后项还有优化的余地。

这就是SVM的特征之一:C越大,就越不能容忍误分类,哪怕margin变得很小也在所不惜;C越小,就越看重margin是否大,哪怕分错数据也无所谓。

43楼 qq_34415740 2016-12-07 16:32发表



厉害了

42楼 独霸首席 2016-11-25 11:19发表



关于:

我有个问题:

你说"由于将C设的很大, cost function只剩下后面的那项。采取简化形式,意在说明问题即可,设 $\theta_0=0$,只剩下 θ_1 和 θ_2 ,则cost function $J(\theta)=1/2 \times \|\theta\|^2$ "

我觉得很奇怪,C大,说明前项的权重大,那更应该是优化前项而不是后项。当然我看视频也没理解为什么这么做,莫非只是为了说明原理,而和C并没有关系?

Re: stellar1 2017-02-17 09:45发表



引用"qq_20371377"的评论:

关于:

我有个问题:

你说"由于将C设的很大, cost function只剩下后面的那项。采取简化形..."

不是不优化前项,而是正因为前项太重要了,所以干脆把前项直接写成限制条件(就是s.t.的部分)。因为当C极大时(极端一点,就比如C是正无穷吧),此时只要有哪怕一个点的cost的值不是0(即这个点没有落在正确分类那一侧的margin之外),那前项就是正无穷。此时无论怎么优化后项都完全没有意义了。要避免这种情况,就必须保证前项中所有的cost都是0,而这就其实已经不是“优化目标”而是“限制条件”了,所以才把前项直接写成了s.t.,只剩下后项还有优化的余地。

这就是SVM的特征之一:C越大,就越不能容忍误分类,哪怕margin变得很小也在所不惜;C越小,就越看重margin是否大,哪怕分错数据也无所谓。

41楼 anjing686 2016-11-19 15:25发表



大神,赞一个,学习了

40楼 Artprog 2016-05-22 15:56发表



我想问个问题啊,为什么C很大,我们就想要方括号里面的误差为零啊?一直不懂,这是里面有什么逻辑关系吗?虽然我们的目的确实是要求最小化误差,但是为什么C大了,方括号里面就必须为0呢(或者说趋近于)?

39楼 白马负金羁 2015-11-24 15:07发表



感谢分享!赞一个!(*^__^*)'

38楼 zhangcsndnet 2015-09-15 12:15发表



楼主您好,前面讲SVM costfunction 的时候 要求 $\theta^T x \geq 1$,后面讲到核函数了就变成了 要求 $\theta^T x \geq 0$ 来判断 $y=1$.这一点我很不解,还请帮忙指点一下。

Re: WangBo_NLPR 2015-12-03 14:07发表



回复zhangcsndnet:引进核函数以后其实应该是: $\theta^T f \geq 1$,这是SVM对分类更严格的体现,可能视频里笔误

37楼 cswwwp 2015-04-23 18:14发表

-  好难。。。
- 36楼 bengion 2015-01-06 21:52发表
 不知道博主学了多长时间,我今年刚开始入手机器学习,数学专业的,还请多多指教,是不是隔行如隔山呢?
- 35楼 某小皮 2014-09-12 22:09发表
 dear小学姐
r=sqrt(rand(100,1));%generate 100 random radius
这个地方为什么要加一个sqrt?去掉不可以吗?
- 34楼 cjc雪狼 2014-07-02 17:03发表
 不错,感谢!
- 33楼 净华 2014-06-24 16:45发表
 其实按照我的理解,SVM的目标主要是在几何间隔最大化,而几何间隔的最大化的目标,就是求出一个超平面w,在该法方向上能将数据分隔的最好,也就是说数据在这个方向上可以整体得到最大坐标值,如何得到最大间隔,加减最小化该法方向的单位长度即 $1/\|w\|$
- 32楼 PassionWei 2014-06-03 10:48发表
 博主,你好,对于初学SVM的,有没有较好的入门材料推荐一下,我只是用来分类的。谢谢
- 31楼 ltt19900822 2014-01-17 20:24发表
 Hello~感觉你对这个算法的理解很到位,我最近有在关注斯坦福公开课,然后觉得这些算法放到实际应用中才有意义,所以,你是用到的时候再看这些算法,还是先看算法,对这些算法都有一个大致的了解之后,遇到问题再用这些算法啊?
- 30楼 LOUIS_liu_ 2013-12-01 20:19发表
 Hi Rachel,请教一下,我现在要做一个,语义分析——情感分析的项目。请问,有什么好的建议。我现在比较迷茫,O(n_n)谢谢!
- 29楼 lkydedeskmate 2013-09-06 21:36发表
 最近一直在看楼主的文章,感觉有很大的收获。
另外想请教一个问题,对于SVM分类算法,当待测样本中出现训练集中不存在的种类(比如训练集的样本都是种类1,2,3,4,而待测样本属于未知的种类5)时,能否将分类输出为“种类未识别”,而不是错误的分到1-4类别中。
相关的研究之前有没有人做过呢?可以用什么手段?如果有,在libsvm工具箱上能否实现呢?
谢谢!
- 28楼 南岳怀让 2013-07-31 09:58发表
 额,有一个比较简单的问题想问一下(一直没想明白),w为什么是垂直于超平面的啊??
- Re: 金咏 2013-11-05 16:17发表
 回复qingyuanxingsi:w是平面的法向量
- 27楼 星夜落尘 2013-07-23 17:54发表
 为什么这个和我在网易公开课看到的不一样呢?
- Re: leasl_Zh 2013-08-27 18:34发表
 回复ADF1230:这是coursera上面的课程,感觉跟斯坦福的公开课比要更浅显一点
- 26楼 pengalg 2013-07-21 19:39发表
 我最近再看stanford公开课,里面讲到L1 Norm Soft Margin,我想问一下,Guass Kernel相当于映射到无穷维,在这个情况下,所有样本数据都是线性可分的,那L1 Norm Soft Margin就失去作用了吧??我理解,L1 Norm Soft Margin是处理线性不可分的情况,如果采用Guass Kernel,在高维空间,就不会存在线性不可分了。。
- 25楼 pengalg 2013-07-21 19:37发表
 fsdf
- 24楼 徘徊者GHB 2013-07-18 09:19发表
 博主你好,经常看到SVM的讲解中提到要将低维特征映射到高维空间中,然后找出最大间隔分类平面。关于这个低维向高维映射这个问题该怎么理解?在上文中好像没太看出来,谢谢

Re: epleone 2014-07-24 13:21发表



回复T800GHB:就是通过核函数

23楼 YangYang_Neng 2013-06-05 16:21发表



你好,有个问题想要请教LZ一下,(文中第一节)logistic的损失函数变成了SVM的损失函数,那么相应于logistic的假设函数形式,SVM的假设函数形式是什么样的?

22楼 haoaina521 2013-05-13 20:45发表



你好,这句话“我们已知boundary的角度和θ向量呈的是90°角(自己画一下就知道了)。”中的90°是怎么确定的???

Re: 天泽28 2016-05-10 09:40发表



回复haoaina521:超平面的法向量和平面垂直

Re: Felsen 2013-06-18 14:20发表



回复haoaina521:我也不知道,同问

Re: 天泽28 2016-05-10 09:41发表



回复jj12345jj198999:超平面的法向量和平面垂直

21楼 小熊不去实验室 2013-04-24 21:33发表



受益匪浅~ 博主完全可以建个ML专栏~

20楼 wodownload2 2013-01-16 10:13发表



你最好把lib-svm的源代码分析下,写个报告,让我们学习下。吼吼

19楼 a120471 2013-01-10 19:15发表



T T 没看懂 我就是想知道怎么写一个基于高斯核函数的分类器。。。老师讲的基本没听明白...课件还特别不好理解。。。看了几篇csdn大牛们写的文章 但是都好多啊。。。公式们看起来太烦了。。看不进去。。怎么办T T 有没有入门一点的文章求..推..荐..

Re: a120471 2013-01-10 19:19发表



回复a120471:BTW 我处于那种对线性可分的SVM稍稍有点理解的程度...甚至还不清楚核函数到底是干什么的...

Re: a120471 2013-01-10 21:07发表



回复a120471:poly核函数一定能把线性不可分变成线性可分的吗?是不是因为它产生了所有组合的情况~

Re: a120471 2013-01-12 18:38发表



回复a120471:怎么最后我还是觉得各种版本的代码都变成二次规划求解了呢 对偶性在实际中用的到么...

18楼 quant_wrz 2013-01-06 16:50发表



博主您好,请教个问题,把feature x化成landmark的目的是为了什么?能说具体点吗?看不太明白

今天突然开不开coursera上的视频相当郁闷,你之前学的时候也会有看不了的情况吗

Re: Rachel-Zhang 2013-01-06 20:49发表



回复quant_wrz:为什么选择landmark呢?因为高斯核函数的方法需要几个kernel,而kernel中有landmark,这里和KNN有些类似。不过KNN是看一个点的N个近邻标号,Kernel的方法是看该点和哪个kernel最近。

17楼 dandie518 2013-01-04 14:40发表



这里关于线性SVM里C的描述,有点问题... C很大并不是形成Large Margin的原因。C约束的是经验风险ER, Large Margin对应SR,所以C过大的时候才会Overfit

Re: Rachel-Zhang 2013-01-06 21:06发表



回复dandie518:and... I'm headache about your abbreviation, can you tell me the meaning of ER & SR?

Re: dandie518 2013-01-08 08:28发表



回复abcjennifer: sorry,我只是路过,没写的很仔细:经验风险 Empirical Risk
SR: Structural Risk, 对应泛化能力,体现在 $\|w\|^2$ 这一项

Re: Rachel-Zhang 2013-01-09 00:42发表



回复dandie518:ok, thank you.

Re: Rachel-Zhang 2013-01-06 21:03发表



回复dandie518:这里是一种解释, 如果用 $y=wx+b$ 表示SVM的boundary的话, objective function为 $1/2\|w\|^2$, s.t. $(wx+b)>=1$ 也是可以的。这两种道理是一样的吧?

Re: dandie518 2013-01-08 09:06发表



回复abcjennifer:"C大, decision boundary则呈现为黑线;如果C不是很大, 就呈现粉线;"

你举的这个例子, 不知道你有没有拿svm算法跑跑看?

按常理来说, C很大的时候svm不愿犯错而牺牲Margin才会得到Margin较小的粉线, 而C不是太大的时候, svm照顾Margin而不惜犯错, 会得到Margin较大的黑线。

从这个意义上也可以看到,C 并不是Largest Margin的原因, 在一定程度上甚至还有矛盾。

btw: 你的文字结论和你画的图是反的。。。

Re: Rachel-Zhang 2013-01-09 00:30发表



回复dandie518:嗯, 这里确实写反了, 已经改过来了。是的, C的选择就是一个分类和margin的tradeoff。为了防止过拟合要让margin大, 所以不能让C过大, 为了防止犯错不能让C过小, 导致前面一项可以无穷大。您说的很对。

Re: dandie518 2013-01-08 08:34发表



回复abcjennifer:svm是有很多个角度可以诠释, 不过你写的这段话恐怕不太妥当——“让C很大, 是一个明智的选择。那么C取一个很大的值会带来什么效果呢?就是我们开头说的结论——SVM是一个large margin classifier”

svm是在保证尽量不错的条件下(几乎以往所有的算法都这么干)追求结构风险的最小(这是svm的核心)。而这个C的意义是在于控制“尽量不出错”, 并不是导致Largest Margin的直接原因。

实际应用中也可以看出这一点, 对C的取法, 是从小开始取, 取到错误的程度可以接受, 就不再增大, 换句话说, 其实是希望C不要太大大...

Re: Rachel-Zhang 2013-01-09 00:41发表



回复dandie518:这里我也比较纠结, 这篇内容是完全按照公开课里Andrew讲的来写的(他说C要取一个很大的值, 可是之前我看到的(好像PRML)里是按照prior<margin大的classifier比较好>来的, 这样只需要最小化 $\|w\|$ (也就是theta), 在这种情况下就不能让C很大。

Re: dandie518 2013-01-09 11:07发表



回复abcjennifer:这可能与解释SVM能达到Largest Margin的逻辑套路有关。作为老师都需要推出支持向量的间距是 $2/\|\theta\|$, 然后就可以说为了间距最大, $\|\theta\|$ 得最小, 所以SVM就能达到最大的Margin。

但是这个套路的前提是所有的样本的判决函数值 $p^*\|\theta\|$ 的绝对值 $>=1$ (或者在松弛后满足), 也就是说全都分对了(或者说在松弛后能当作是分对了)

假使不引入C, 显式约束本身就可以满足这个条件, 那自然只需要一味的缩小 $\|\theta\|$ 。但是如果引入了C, 要达到这个前提, C就需要足够的大, 否则这个前提都达不到, 其实SVM的最大Margin也就没了根基。

所以我想作者大概是想提醒读者C在这件事里的积极作用。C的作用颇像中国文化的阴阳学说, C就好像是“阴”的力量, 虽然看似和“阳”(Margin)矛盾, 但是却以制约的方式促成了阳达到自己的极致。

16楼 fangzheng354 2012-12-26 15:21发表



崇拜 好详细啊

15楼 csmri 2012-11-09 10:25发表



其实就是一个核学习再加个正则化, 某种意义上说核 的作用也是正则化

14楼 fsi206914 2012-10-23 02:59发表



你好lz, 为什么logistic regression 的 cost function 的后半部分会有 $(\lambda/2m) \sum(\theta_i^2)$ 后面这一项, 我看前面的章节, 貌似没有也, 这是为什么?

Re: Rachel-Zhang 2012-10-23 09:11发表

 回复fsi206914: 避免过拟合, 这里是引入的regularization项, 参考第三讲<http://blog.csdn.net/abcjennifer/article/details/7716281>中的(十一)、Regularized Logistic Regression

Re: fsi206914 2012-10-23 09:47发表

 回复abcjennifer: Thx. I've Understand now. Have a good day.

13楼 wwy250 2012-10-15 01:55发表

 看了一些关于SVM的论文, 都是讲用于classification或regression
请问SVM可以用于learn p(y|x)么

Re: Rachel-Zhang 2012-10-15 07:58发表

 回复wwy250: 概率是一个统计值, 如果想求 $P(y|x)$ 通常用bayesian吧.....
 $P(y|x) \propto P(x|y)P(y)$

Re: wwy250 2012-10-15 16:36发表

 回复abcjennifer: 这倒是 只是想问问 可不可以通过修改似然度函数改造成SVM 毕竟SVM是结构风险最小化而不是经验风险最小化 并且支持无限维特征值嘛
另外如果这里 y 与 x 均不为离散量 是不是只有假设 y 服从关于 x 的什么什么分布 然后根据最大似然解参数呢

Re: Rachel-Zhang 2012-10-15 19:29发表

 回复wwy250: 第一个问题, SVM的宗旨是maximize支持向量的margin, 加入概率这个。。。没想清楚你想在哪里加呢?
第二个问题, 是的~ 如果按Bayesian Decision Theory概率模型来, 先假设一个分布(比如高斯), 然后 $\theta = \text{maximize } D(\theta)$ 的那个 θ , (D 是全部数据)也就是最大似然估计; 如果看做回归问题, 可以进行线性/非线性回归, 定义loss function, 通过minimize loss function求解参数。

12楼 LarryNLPiR 2012-10-13 11:10发表

 楼主很有眼光啊, coursera上面的课程都很经典, 而且很适合自学, 做得非常人性化。

Re: Rachel-Zhang 2012-10-13 16:17发表

 回复yangliuy: 呵呵~是滴~coursera上的课程都很好

Re: LarryNLPiR 2012-10-13 16:22发表

 回复abcjennifer: 博主做CV? 浙大图形学很厉害啊, 研一估计要上课吧, 这个系列写的挺好的, 期待更多精彩博文

Re: Rachel-Zhang 2012-10-13 16:46发表

 回复yangliuy: 呃~以前做cv的~呵呵, 现在课很多, 没什么时间写了, 我尽量~看了你的blog, 内容很多啊~在做数据挖掘?

Re: LarryNLPiR 2012-10-13 17:00发表

 回复abcjennifer: 对啊, 偏文本挖掘和NLP, 也很关注机器学习问题, 你的博客人气真旺啊, 呵呵

11楼 三千分之一的爱 2012-08-21 21:38发表

 这么NB, 完全看不懂啊

10楼 likun_tech 2012-08-19 16:41发表

 你是用什么语言或者工具做实验的?

Re: Rachel-Zhang 2012-08-22 22:53发表

 回复likun_tech: matlab

9楼 星夜落尘 2012-08-17 13:54发表

 毕业设计做的SVM/SVR, 这东西还是还挺不错的, 可惜效率太低了, 在没有更好的算法出现之前, 这个只能和ANN一样摆在书桌上。

8楼 wait_rabbit 2012-08-17 00:46发表

 我对machine learning非常感兴趣, 可惜不是学这个专业的。博主可否介绍一下, 学习这门技术, 所必须的数学基础知识有哪些? 谢谢!

Re: Rachel-Zhang 2012-08-17 09:29发表



回复wait_rabbit:数值分析, 比如线性代数和概率分析是基础内容:)

7楼 椰树海岛 2012-08-16 09:02发表



神,码

6楼 牧风 2012-08-15 16:32发表



整体味道很不错

缺点就是有点生搬PPT了

做参数寻优的话, matlab本身就有带GA工具箱, 可以试试, 很好用~
不过这几年小蜜蜂算法比较火~

楼主加油呀

研究数据挖掘是我的遗志.....

Re: Rachel-Zhang 2012-08-15 20:41发表



回复alex0:啊哈, 谢谢!话说, 小蜜蜂是什么?

Re: 牧风 2012-08-16 15:21发表



回复abcjennifer:最主要的是Artificial bee colony, 但还有非常多的分支

5楼 happynear 2012-08-15 16:18发表



啊, 我也是即将进入研一的学生, 方向也是CV。但跟你相比, 我还差得很远。你这几篇文章都研究得好深入, 我之前看这些方法都是浅尝辄止, 以后要以你为榜样, 认真去学习了。

Re: Rachel-Zhang 2012-08-15 20:42发表



回复happynear:谢谢哈~ 过奖了, 东西写出来理解会更深~

4楼 Emiyasstar__ 2012-08-14 19:17发表



libsvm有个国内的扩充版本, 加入了ga,pso等自适应寻优算法, faturo写的, libsvm-mat-2.89-3[FarutoUltimate3.0]搜搜这个

Re: Rachel-Zhang 2012-08-15 20:43发表



回复Emiyasstar__:好的~谢谢!

3楼 sh1t2 2012-08-14 10:08发表



感觉楼主每天都有时间学习新东西啊, 项目压力和科研压力不大吗?这种自由研究的感觉真好。
ps:楼主硕士第几年了?做视觉方向毕业后有何打算?给我们后来人参考参考~

Re: Rachel-Zhang 2012-08-14 10:16发表



回复sh1t2:呵呵~我在享受大四最后一个暑假~
嗯, 以后想做cv.....

Re: Garfield-Dong 2014-05-22 16:07发表



回复abcjennifer:Rachel你太幸福了, 我大四毕业七天后就来给老师干项目干到了现在, 至今每年寒暑假各15天。。。至今机器学习没看完。
至今还在带着师弟调代码。。羡慕你。。。你是几年的硕士?毕业神马打算?

Re: Rachel-Zhang 2014-05-24 10:49发表



回复quan1614:辛苦啊.....我3年的还差一年毕业, 毕业干什么看实习情况啦(*^__^*)

2楼 pl__ 2012-08-14 09:48发表



我记得有个脚本能自求优化参数的, 关于libsvm

Re: Rachel-Zhang 2012-08-14 10:15发表



回复pl__:哦?什么脚本还记得么?给个提示

Re: pl__ 2012-08-14 15:39发表



回复abcjennifer:我找找 好久之前听别人说过

四、准备训练参数

因为svm训练需要手动调整参数,一般采用默认的情况即可,这里还是为了体验libsvm,可以用libsvm自带的grid.py来自动暴力搜索最好的参数c和g,c表示惩罚系数,g表示gamma系数。
<http://www.cnblogs.com/tornadomeet/archive/2012/04/12/2443993.html>

Re: [Rachel-Zhang](#) 2012-08-14 18:07发表

回复pl___:grid.py ~ thx

1楼 [puretech_](#) 2012-08-12 21:16发表

看不懂, mark之。

您还没有登录,请[\[登录\]](#)或[\[注册\]](#)

* 以上用户言论只代表其个人观点, 不代表CSDN网站的观点或立场

[公司简介](#) | [招贤纳士](#) | [广告服务](#) | [联系方式](#) | [版权声明](#) | [法律顾问](#) | [问题报告](#) | [合作伙伴](#) | [论坛反馈](#)

网站客服 | 杂志客服 | 微博客服 | webmaster@csdn.net | 400-660-0108 | 北京创新乐知信息技术有限公司 版权所有 | 江苏知之为计算机有限公司 |

江苏乐知网络技术有限公司

京 ICP 证 09002463 号 | Copyright © 1999-2017, CSDN.NET, All Rights Reserved 