1.LR，SVM

1.LR

4. 逻辑回归 (Logstic Regression, LR)

本质上是线性回归，只是在结果前加了一层逻辑函数。

即：先把特征线性求和，然后使用函数 $g(z)$ 作为假设函数来预测。

$g(z)$ 为sigmoid $\Rightarrow g(z)=\frac{1}{1+e^{-z}}$

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

二分类的损失函数 (交叉熵) | 对数似然损失函数

$$L=\frac{1}{m}\sum_{i=1}^{m} l(\hat{g}^{(i)}, y^{(i)})=\frac{1}{m}\sum_{i=1}^{m}\left(-y^{(i)}\log(\hat{g}^{(i)})-(1-y^{(i)})\log(1-\hat{g}^{(i)})\right)$$

逻辑回归实现多分类：

使用softmax构造模型解决多分类。输出的值对应于每个类别的概率。

2.SVM二分类

2.1原理

Date:　　Page:

5. SVM (Support Vector Machine)　　(原理)

5.1 用于解决二分类或多分类问题。

SVM的目标是寻找一个最优化超平面在空间中分割两类数据，这个最优化超平面需要满足的条件是：离其最近的点到其距离最大化。而这些点称为支持向量。

2.2 公式推导

5.2 • 划分超平面可通过如下线性方程来描述：

$$w^Tx+b=0$$

样本空间中任意点x到超平面(w,b)的距离可写为：

$$r=\frac{|w^Tx+b|}{||w||}$$

假设超平面(w,b)能将训练样本正确分类，有

$$\begin{cases} w^Tx_i+b \geq +1, & y_i=+1 \\ w^Tx_i+b \leq -1, & y_i=-1 \end{cases} \quad (1)$$

• 两个异类支持向量到超平面的距离之和为：（间隔 margin）

$$r=\frac{2}{||w||} \quad (2)$$

• 欲找到具有"最大间隔"(maximum margin)的划分超平面，即找满足(1)中的w与b，使r最大，即

$$\max_{w,b} \frac{2}{||w||}$$

$$s.t. \ y_i(w^Tx_i+b) \geq 1.$$

最大化$||w||^{-1}$，等价于最小化$||w||^2$，优化目标为

$$\begin{cases} \min_{w,b} \frac{1}{2}||w||^2 \\ s.t. \ y_i(w^Tx_i+b) \geq 1 \end{cases}$$

接下来为拉格朗日乘子法转为"对偶问题"求解

2.3核函数

5.4核函数

将不可分的数据映射到高维可分 (高斯核，多项式核…)

什么时候线性核，什么时候高斯核？

当数据的特征提取地较好，问题线性可分 则可采用线性核；

若特征较少，样本适中，，问题线性不可分 ⇒高斯核

核方法

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

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

2.4SVM的硬间隔，软间隔

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

$$s.t. \quad y^{(i)}(w^T x^{(i)}+b) \geq 1$$

硬间隔

$$\min_{w,b,\xi} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m} \xi_i$$

$$s.t. \quad y^{(i)}(w^T x^{(i)}+b) \geq 1-\xi_i$$

$$\xi_i \geq 0$$

软间隔

软间隔是在硬间隔的基础之上引入了松弛变量。(为了防止过拟合)

3.SVM回归

做回归：

分类的 hinge loss

$$\begin{cases} \min_{w,b} \frac{1}{2}\|w\|^2 & \text{：让支持向量之间距离更大} \\ s.t. \quad y_i(w_i\,\phi(x_i)+b) \geq 1 & \text{：让样本都正确分类} \end{cases}$$

而对于回归模型：支持向量之间距离大不重要，但约束条件改变。回归模型的目标是让训练集中的每个点 (x_i, y_i) 尽量拟合到一个线性模型 $y_i = w\cdot\phi(x_i)+b$。

然后，定义个常量 $\varepsilon>0$，使 $|y_i - w_i\phi(x_i) - b| - \varepsilon$ 最小

对于某个点 (x_i, y_i)，若 $|y_i - w\cdot\phi(x_i) - b| \leq \varepsilon$，则完全无损失。若 $|y_i - w\cdot\phi(x_i) - b| > \varepsilon$，则对应的损失为 $|y_i - w\cdot\phi(x_i) - b| - \varepsilon$

4.SVM优缺点

优点：1) 面对高维数据，可以使用核函数进行有效处理

2) 使用核函数可以解决非线性分类.

缺点：

1) SVM算法对大规模训练样本难以实施

2) 对缺失数据敏感，对参数和核函数的选择敏感

3) 解决多分类问题困难：

经典的SVM只给出了二分类算法，而在实际应用中，一般要解决多分类问题，可以通过多个支持向量机组合来解决。

5.SVM与LR

1) LR是参数模型，SVM是非参数模型

2) LR采用的损失函数是logisticalloss，SVM是hingeloss.

3) 在学习时，SVM只考虑与分类最相关的少数支持向量点。

LR的模型相对简单，在进行大规模线性分类时比较方便。

6.调参

1)SVM

①分类：from sklearn.svm import SVC

clf = SVC(kernel='linear').fit(X,y)

调1：核函数有"linear"、"rbf"、"poly"(多项式核)

调2：C：惩罚参数：默认设置1。（C↑，过拟合，C↓，欠拟合）

C越大，相当于惩罚松弛变量，希望松弛变量接近0，即对误分类的惩罚增大，趋向于训练集全对，但泛化能力弱（过拟合）

C越小，对误分类的惩罚减小，允许容错，将他们当成噪声点，泛化能力强。

②回归：from sklearn.svm import SVR