Loss in ML
Part 1: Linear Model

Hao Liang 2019/12/6

损失函数(Loss)

- Loss衡量模型中参数的好坏

最小二乘法

- 假设y和x具有线性关系

$$y = kx + b$$

- 测量到的y具有高斯误差，测量的x

$$y \sim \text{gauss}(\mu = {\color{red}w}x + {\color{red}b}, \sigma)$$

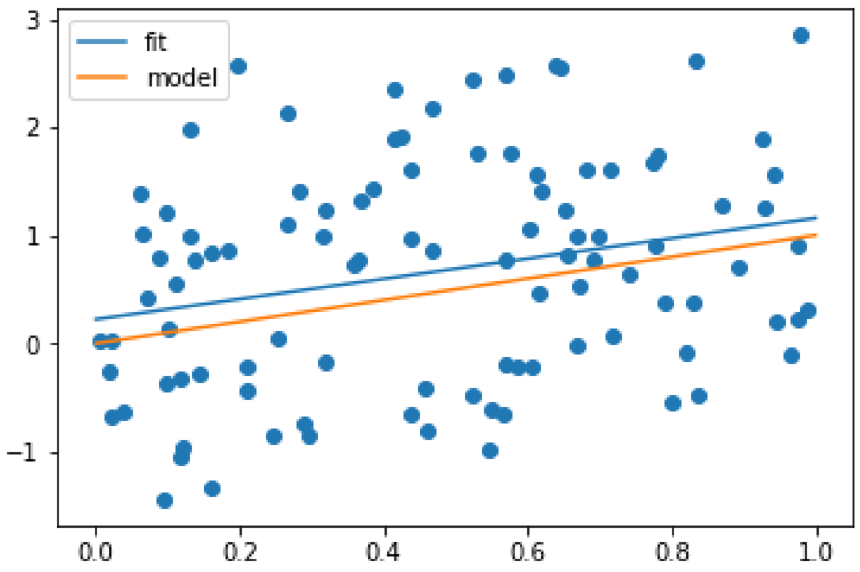
现在我们一堆数据(x_i, y_i)，得到这一堆数据的概率是

$$\Pi_i \frac{1}{\sqrt{2\pi\sigma}} \exp(-(wx_i + b - y_i)^2/2\sigma^2)$$

唠咯一下，取负，再移去常量

$$loss = \sum_i \frac{({\color{red}w}x_i + {\color{red}b} - y_i)^2}{2}$$

更进一步：正则化（岭回归，套索回归和弹性网络），支持向量回归



逻辑回归

- 二分类问题，$x_{\mathrm{i}} \in R, y_i \in \{0,1\}, i \in [1, m]$
- 假定正类$y = 1$的概率p_1是

$$p_1(x) = \frac{1}{1 + \exp(-(w \cdot x + b))}$$

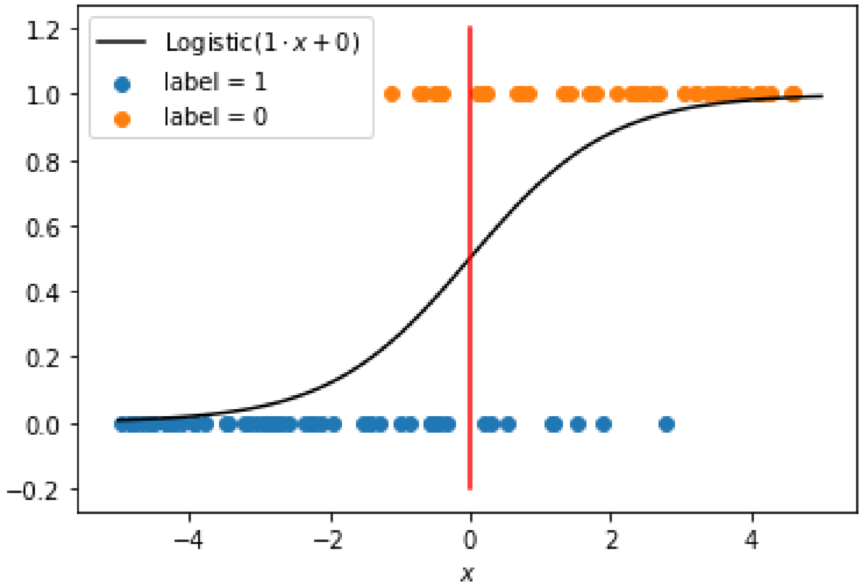
- 假设我们有一套数据$\{\mathrm{x_i}, y_i\}$，那么观测到的概率是

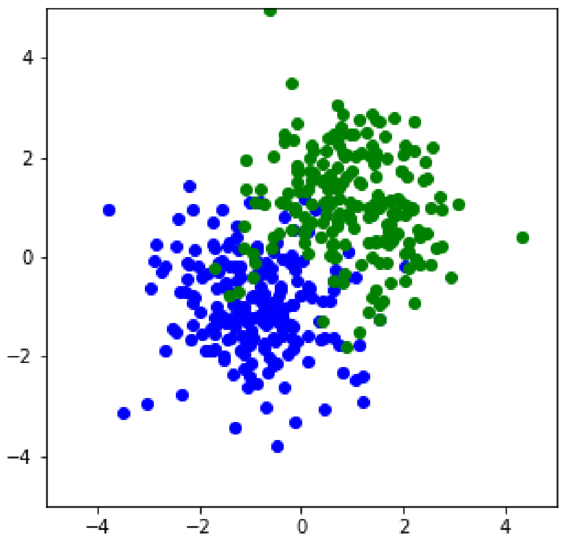
$$\Pi_i p_1(x_i)^{y_i} \cdot (1 - p_1(x_i))^{1-y_i}$$

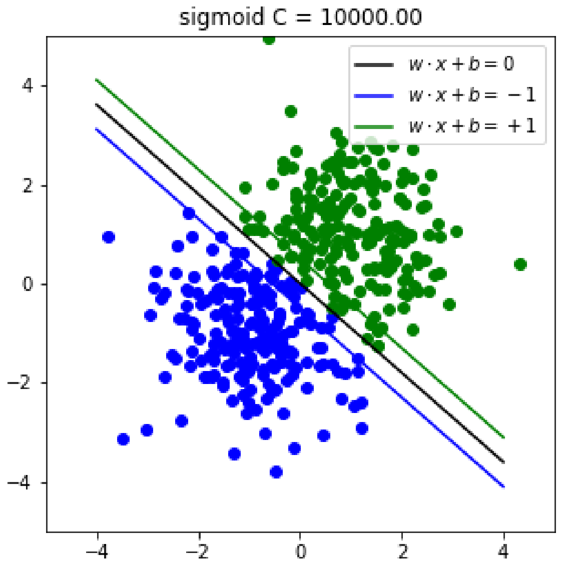
- 唠咯一下，取负，再移去常量

$$loss = -\frac{1}{m}\sum_i y_i \log(p_1(x_i)) + (1 - y_i)\log(1 - p_1(x_i))$$

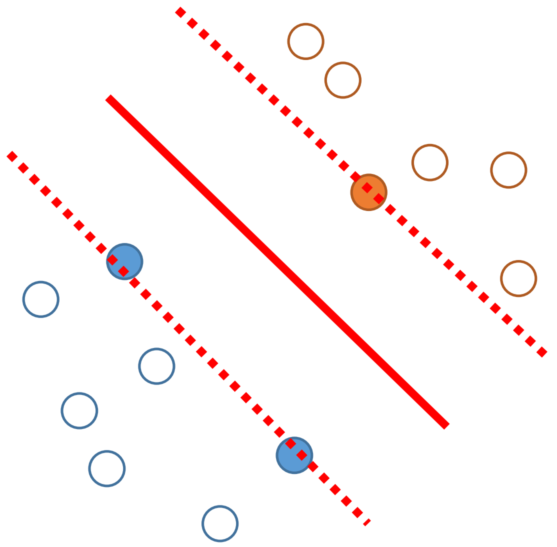
更进一步：多分类问（softmax），正则化（L1和L2），kernel逻辑回归



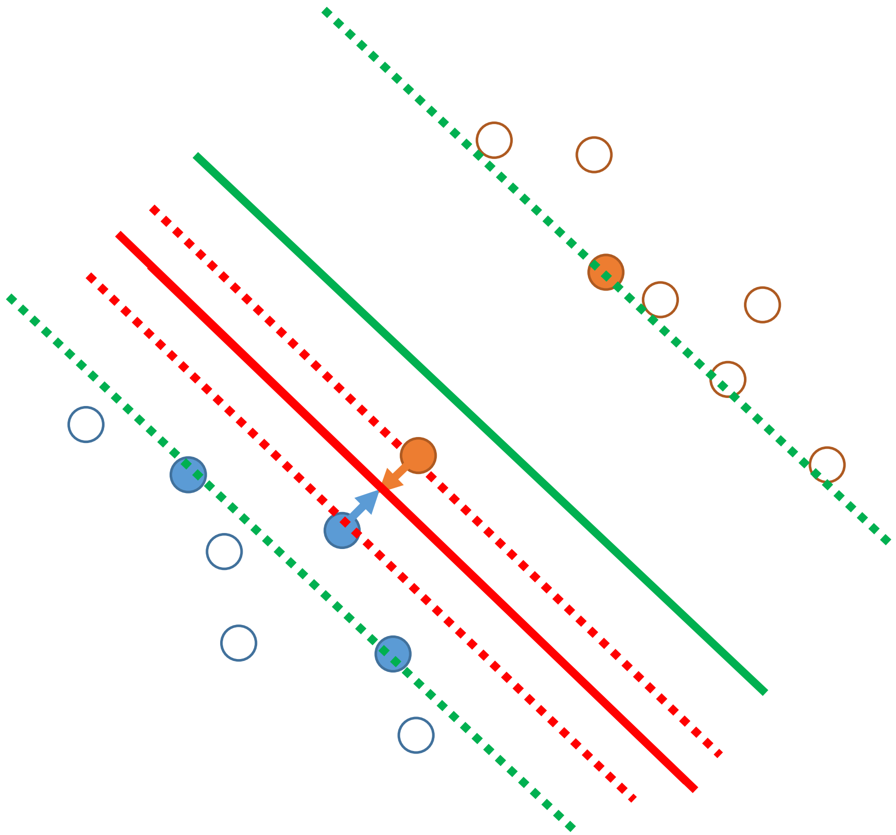




支持向量机



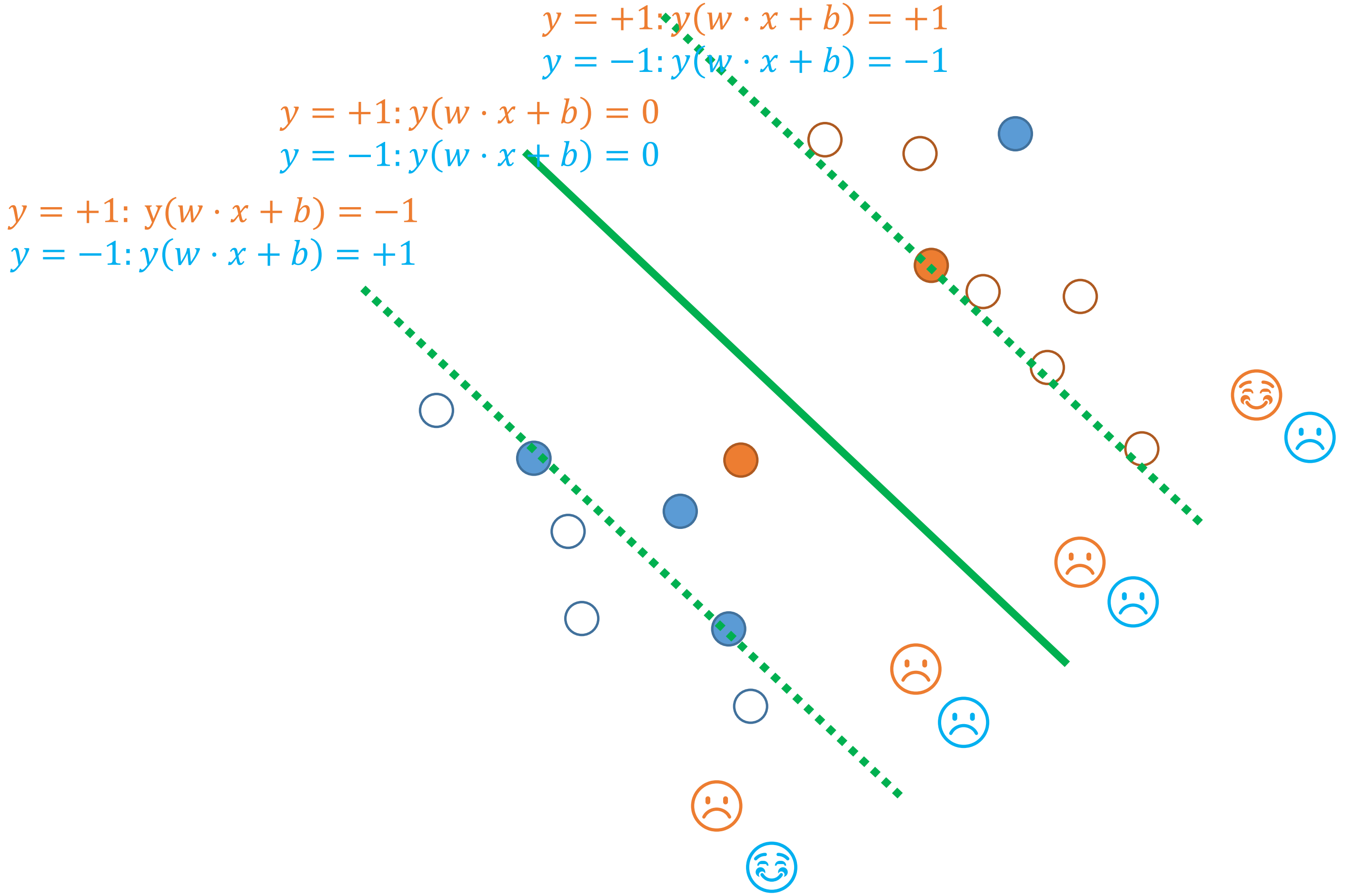
硬间隔算法



$y = 1:$
$w \cdot x + b = 1$

软间隔算法

支持向量机



$$loss = \sum_i \frac{1}{2} w^2 + C \begin{pmatrix} 0 & : y_i(w \cdot x_i + b) \geq 1 \\ 1 - y_i(w \cdot x_i + b) & : y_i(w \cdot x_i + b) < 1 \end{pmatrix}$$

Thanks

AdaBoost & GBDT

- $loss = \sum_i \exp(-y_i \sum_j \alpha_j g_j(x_i))$
- Boosting:改进之前的模型
- $loss = \sum_i l(-y_i(G(x_i) + \alpha g(x_i)))$
- $g(x_i)$修正权重后的学习器的结果

$$g(x_i) = \begin{cases} +1 \text{ 预测为正类} \\ -1 \text{ 预测为负类} \end{cases}$$

- 优化α
- 更新$G(x_i)$

$$G(x_i) \Leftarrow G(x_i) + \alpha g(x_i)$$

- $loss$不限
- Boosting:改进之前的模型
- $loss = \sum_i l(-y_i(G(x_i) + \gamma g(x_i)))$
- $g(x_i)$为负梯度

$$g(x_i) \Leftarrow -\frac{\partial loss}{\partial G(x_i)}\bigg|_{g(y_i)=0}$$

- 优化γ
- 更新$G(x_i)$，但是通常不会走γ那么多: $v < 1$

$$G(x_i) \Leftarrow G(x_i) + v\gamma g(x_i)$$