Logistic Regression

实战目标：

1. 掌握逻辑回归模型的基本原理
2. 实现模型训练、预测、评价、存储和加载
3. 实现利用向量运算的模型训练
4. 可视化

一元逻辑回归

设数据集有 m 个样本 $D=\{(x^{[1]},y^{[1]}),(x^{[2]},y^{[2]}),...,(x^{[m]},y^{[m]})\},x\in\mathbb{R},y\in\{0,1\}$

模型函数：

逻辑回归是在线性回归的基础上增加了sigmoid函数

sigmoid函数：

$$sigmoid(x)=\frac{1}{1+e^{-x}}$$

模型函数为：

$$\hat{y}=sigmoid(a+bx)$$

交叉熵损失函数

$$J(a,b)=-\frac{1}{m}\Sigma_{i=1}^{m}(y^{[i]}*log(\hat{y}^{[i]})+(1-y^{[i]})*log(1-\hat{y}^{[i]}))$$

损失函数在 a, b 方向上的偏导数

1. 令 $y=sigmoid(z)$ 则：

$$\begin{aligned}\frac{d}{dz}sigmoid(z)&=\frac{d}{dz}\frac{1}{1+e^{-z}}\\&=\frac{-(-e^{-z})}{(1+e^{-z})^2}\\&=\frac{e^{-z}}{(1+e^{-z})^2}\\&=\frac{1+e^{-z}-1}{(1+e^{-z})^2}\\&=\frac{1}{1+e^{-z}}-\frac{1}{(1+e^{-z})^2}\\&=y-y^2\\&=y*(1-y)\end{aligned}$$

2. 在 a, b 方向上的偏导数分别为：

$$
\begin{aligned}
a: \frac{dJ(a,b)}{da} &= \frac{d}{da}(-\frac{1}{m}\Sigma_{i=1}^{m}(y^{[i]} * log(\hat{y}^{[i]}) + (1-y^{[i]}) * log(1-\hat{y}^{[i]}))) \\
&= -\frac{1}{m}\Sigma_{i=1}^{m}\frac{d}{da}(y^{[i]} * log(\hat{y}^{[i]}) + (1-y^{[i]}) * log(1-\hat{y}^{[i]}))) \\
&= -\frac{1}{m}\Sigma_{i=1}^{m}(y^{[i]} * \frac{d}{da}log(\hat{y}^{[i]}) + (1-y^{[i]}) * \frac{d}{da}log(1-\hat{y}^{[i]})) \\
&= -\frac{1}{m}\Sigma_{i=1}^{m}(y^{[i]} * \frac{1}{\hat{y}^{[i]}} * \hat{y}^{[i]} * (1-\hat{y}^{[i]}) * 1 + \quad (1-y^{[i]}) * \frac{-1}{1-\hat{y}^{[i]}} * (1-\hat{y}^{[i]}) * \hat{y}^{[i]} * 1) \\
&= -\frac{1}{m}\Sigma_{i=1}^{m}(y^{[i]} * (1-\hat{y}^{[i]}) - (1-y^{[i]}) * \hat{y}^{[i]}) \\
&= -\frac{1}{m}\Sigma_{i=1}^{m}(y^{[i]} - y^{[i]} * \hat{y}^{[i]} - \hat{y}^{[i]} + y^{[i]} * \hat{y}^{[i]}) \\
&= -\frac{1}{m}\Sigma_{i=1}^{m}(y^{[i]} - \hat{y}^{[i]}) \\
&= \frac{1}{m}\Sigma_{i=1}^{m}(\hat{y}^{[i]} - y^{[i]}) \\
b: \frac{dJ(a,b)}{db} &= \frac{d}{db}(-\frac{1}{m}\Sigma_{i=1}^{m}(y^{[i]} * log(\hat{y}^{[i]}) + (1-y^{[i]}) * log(1-\hat{y}^{[i]}))) \\
&= -\frac{1}{m}\Sigma_{i=1}^{m}\frac{d}{db}(y^{[i]} * log(\hat{y}^{[i]}) + (1-y^{[i]}) * log(1-\hat{y}^{[i]}))) \\
&= -\frac{1}{m}\Sigma_{i=1}^{m}(y^{[i]} * \frac{d}{db}log(\hat{y}^{[i]}) + (1-y^{[i]}) * \frac{d}{db}log(1-\hat{y}^{[i]})) \\
&= -\frac{1}{m}\Sigma_{i=1}^{m}(y^{[i]} * \frac{1}{\hat{y}^{[i]}} * \hat{y}^{[i]} * (1-\hat{y}^{[i]}) * x^{[i]} + \quad (1-y^{[i]}) * \frac{-1}{1-\hat{y}^{[i]}} * (1-\hat{y}^{[i]}) * \hat{y}^{[i]} * x^{[i]}) \\
&= -\frac{1}{m}\Sigma_{i=1}^{m}(y^{[i]} * (1-\hat{y}^{[i]}) - (1-y^{[i]}) * \hat{y}^{[i]}) * x^{[i]} \\
&= -\frac{1}{m}\Sigma_{i=1}^{m}(y^{[i]} - y^{[i]} * \hat{y}^{[i]} - \hat{y}^{[i]} + y^{[i]} * \hat{y}^{[i]}) * x^{[i]} \\
&= -\frac{1}{m}\Sigma_{i=1}^{m}(y^{[i]} - \hat{y}^{[i]}) * x^{[i]} \\
&= \frac{1}{m}\Sigma_{i=1}^{m}(\hat{y}^{[i]} - y^{[i]}) * x^{[i]}
\end{aligned}
$$

梯度下降法求解:

参数延着逆梯度方向下降，逐步逼近损失函数的极小值

设学习率为 η:

$$
\begin{aligned}
a = a - \eta * \frac{dJ(a,b)}{da} \\
b = b - \eta * \frac{dJ(a,b)}{db}
\end{aligned}
$$

写成向量:

令$X = [x^{[1]}, x^{[2]}, ..., x^{[m]}], Y = [y^{[1]}, y^{[2]}, ..., y^{[m]}], X \in \mathbb{R}^m, Y \in \{0,1\}^m$

损失函数在 a, b 方向上的偏导数:

$$
\begin{aligned}
a: \frac{dJ(a,b)}{da} &= \frac{1}{m}(\hat{Y} - Y) \\
b: \frac{dJ(a,b)}{db} &= \frac{1}{m}(\hat{Y} - Y) \cdot X
\end{aligned}
$$

梯度下降:

同上

多元线性回归

设数据集有 m 个样本
$D=\{(X^{[1]},y^{[1]}),(X^{[2]},y^{[2]}),...,(X^{[m]},y^{[m]})\},X\in\mathbb{R}^n,y\in\{0,1\}$

模型函数:

$$\hat{y}=sigmoid(\theta_0+\theta_1x_1+\theta_2x_2+...+\theta_nx_n)$$

为了方便计算，对 X 再添加一维，即令 $x_0=1,X\in\mathbb{R}^{n+1}$, 令$\Theta=[\theta_0,\theta_1,...,\theta_n]^T,\Theta\in\mathbb{R}^{n+1}$, 则:

$$\begin{aligned}\hat{y}&=sigmoid(\theta_0x_0+\theta_1x_1+\theta_2x_2+...+\theta_nx_n)\\&=sigmoid(\Theta^TX)\end{aligned}$$

交叉熵损失函数

$$J(a,b)=-\frac{1}{m}\Sigma_{i=1}^m(y^{[i]}*log(\hat{y}^{[i]})+(1-y^{[i]})*log(1-\hat{y}^{[i]}))$$

损失函数在 θ_k 上的偏导数:

$$\frac{dJ(\Theta)}{d\theta_k}=\frac{1}{m}\Sigma_{i=1}^m(\hat{y}^{[i]}-y^{[i]})*X_k^{[i]}$$

梯度下降法求解:

参数延着逆梯度方向下降，逐步逼近损失函数的极小值

设学习率为 η，更新θ_k:

$$\theta_k=\theta_k-\eta*\frac{dJ(\Theta)}{d\theta_k}$$

更有效的表示

损失函数在 Θ 上的偏导数:

$$\frac{dJ(\Theta)}{d\Theta}=\frac{1}{m}\Sigma_{i=1}^m(\hat{y}^{[i]}-y^{[i]})*X^{[i]}$$

梯度下降法求解:

依旧设学习率为 η， 此时的梯度下降变化为

$$\Theta=\Theta-\eta*\frac{dJ(\Theta)}{d\Theta}$$