1. Logistic回归

<1> $\hat{y}^{(i)} = \sigma(w^T x^{(i)} + b)$, $\sigma(z^{(i)}) = \frac{1}{1+e^{-z^{(i)}}}$

损失函数: $L(\hat{y}, y) = -[y\log\hat{y} + (1-y)\log(1-\hat{y})]$, $\hat{y} = \sigma(w^T x^{(i)} + b)$

推导: 假设 $P(y=1|x;\theta) = h_\theta(x)$, $P(y=0|x;\theta) = 1 - h_\theta(x)$

即 $P(y|x;\theta) = [h_\theta(x)]^y [1-h_\theta(x)]^{1-y}$

又设m个样本是独立生成的, 则似然函数为 $L(\theta) = P(\vec{y}|x;\theta) = \prod_{i=1}^{m} [h_\theta(x^{(i)})]^{y^{(i)}} [1-h_\theta(x^{(i)})]^{1-y^{(i)}}$

取对数: $l(\theta) = \log L(\theta) = \sum_{i=1}^{m} y^{(i)} \log h(x^{(i)}) + (1-y^{(i)}) \log(1-h(x^{(i)}))$

⇒要使极大似然函数最大, 则⇒ $\max l(\theta)$

而代价函数是一个要取极小值的式子, 则在似然对数函数前取负号.

代价函数: 取m平均值: $J(w,b) = -\frac{1}{m} \sum_{i=1}^{m} y^{(i)} \log h(x^{(i)}) + (1-y^{(i)}) \log(1-h(x^{(i)}))$

$= \frac{1}{m} \sum_{i=1}^{m} L(\hat{y}^{(i)}, y^{(i)})$

⇒目标: 解出使 $J(w,b)$ 最小化的 w, b.

<2> 梯度下降步骤:

重复直至收敛:

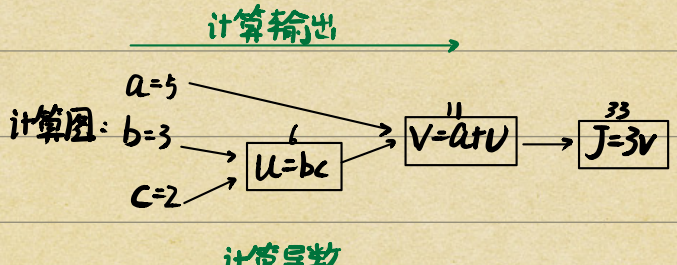
$$w := w - \alpha \frac{dJ(w,b)}{dw}$$ → 变量名: dw

$$b := b - \alpha \frac{dJ(w,b)}{db}$$ → 变量名: db

<3> 计算图

例: 设 $J(a,b,c) = 3(a+bc)$.

∴ 计算顺序: $u = bc$
$V = a+u$
$J = 3V$.



代码中, 写反向传播时, 变量名应设为:

$\frac{d\text{最终输出变量}}{d\text{变量}} = dJ\,dvar \Rightarrow dvar$

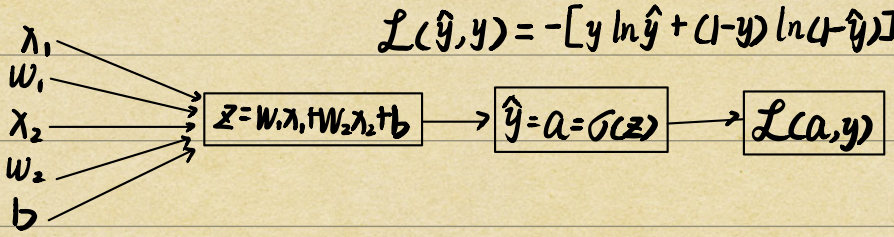
$\frac{dJ}{dv}$ ⇒一步反向传播 $\frac{dJ}{da} = \frac{dJ}{dv} \cdot \frac{dv}{da}$ ←两步反向传播

∴dvar即最终输出变量对它的导数.

例如: $\frac{dJ}{dv} \Rightarrow dv$; $\frac{dJ}{da} \Rightarrow da$

⇒ Logistic Regression Gradient Descent

$$\mathcal{L}(\hat{y},y) = -[y\ln\hat{y} + (1-y)\ln(1-\hat{y})]$$



任务：修改 w 与 b，来降低 $\mathcal{L}$.

计算：

$$\begin{cases} da = -\frac{y}{a} + \frac{1-y}{1-a} \\ dz = \frac{d\mathcal{L}}{da}\cdot\frac{da}{dz} = \frac{d\mathcal{L}}{da}\cdot g'(z) = a - y \quad \text{推：} \frac{da}{dz} = \frac{e^{-z}}{(1+e^{-z})^2} = \frac{1}{1+e^{-z}} - \frac{1}{(1+e^{-z})^2} = a(1-a) \\ dw_i = x_i \cdot dz \\ db = dz \end{cases}$$

反向传播：计算 dw_i 和 db，然后 $\begin{cases} w_i := w - \alpha\, dw_i \\ b := b - \alpha \cdot db \end{cases}$

⇒ 当有 m 个样本时：

代价函数：$J(w,b) = -\frac{1}{m}\sum_{i=1}^{m} y^{(i)}\log h(x^{(i)}) + (1-y^{(i)})\log(1-h(x^{(i)}))$

$$\Rightarrow \frac{\partial J(w,b)}{\partial w_i} = \frac{1}{m}\sum_{i=1}^{m}\frac{\partial}{\partial w_i}\mathcal{L}(a^{(i)}, y^{(i)})$$

初始化：$J=0, dw_1=0, dw_2=0, db=0$

for $i=1$ to m:

$z^{(i)} = w^T x^{(i)} + b$

$a^{(i)} = \sigma(z^{(i)})$

} 计算各样本的激活函数

正向传播 ⟶ 为反向传播作准备

$J \mathrel{+}= -[y^{(i)}\log a^{(i)} + (1-y^{(i)})\log(1-h(x^{(i)}))]$ ⇒ 计算初始 w 下的 cost，用于检测收敛效

$dz^{(i)} = a^{(i)} - y^{(i)}$

$dw_1 \mathrel{+}= x_1^{(i)} dz^{(i)}$, $dw_2 \mathrel{+}= x_2^{(i)} dz^{(i)}$

$db \mathrel{+}= dz^{(i)}$

} 反向传播：

计算各级导数值

$[y^1\ y^2 \cdots]$

$a \Rightarrow [a^1\ a^2 \cdots]$

$J /= m$, $dw_1 /= m$, $dw_2 /= m$, $db /= m$

$w_1 = w_1 - \alpha\, dw_1$

$w_2 = w_2 - \alpha\, dw_2$

$b = b - \alpha\, db$

} 梯度下降

<4> 向量化 $z = np.dot(w, x) + b$; 初始化 dw: $dw = np.zeros(n_x, 1)$

$$X = \begin{bmatrix} | & | & & | \\ x^{(1)} & x^{(2)} & \cdots & x^{(m)} \\ | & | & & | \end{bmatrix}_{(n_x, m)}$$

$\therefore Z = w^T X + b$ (numpy 的广播) 即 $Z = w^T X + [b\ b \cdots b]$ $Z = np.dot(w.T, X) + b$ ↱ w的转置

$A = [a^{(1)}\ a^{(2)} \cdots a^{(m)}] = \sigma(Z)$

$dw = \frac{1}{m}\sum_{i=1}^{m} x^{(i)} dz^{(i)}$ $db = \frac{1}{m}\sum_{i=1}^{m} dz^{(i)}$ $db = \frac{1}{m} \times np.sum(dz)$, $dw = \frac{1}{m} \times np.dot(X, dz.T)$

$dw = \frac{1}{m} X \cdot dz^T \longrightarrow \frac{1}{m}\begin{bmatrix} x^{(1)} & x^{(2)} & \cdots \end{bmatrix}\begin{bmatrix} dz^{(1)} \\ dz^{(2)} \\ \vdots \end{bmatrix} = \frac{1}{m}\left[x^{(1)} dz^{(1)} + x^{(2)} dz^{(2)} + \cdots\right]$

$= \frac{1}{m}\left[dw^{(1)} + dw^{(2)} + \cdots\right] = \frac{1}{m}\begin{bmatrix} dw_1^{(1)} + dw_1^{(2)} + \cdots \\ dw_2^{(1)} + dw_2^{(2)} + \cdots \end{bmatrix} = \frac{1}{m}\begin{bmatrix} z.dw_1 \\ z.dw_2 \\ \vdots \end{bmatrix}_{n_x \times 1}$

向量化实现 Logistic Regression:

$z = np.dot(w.T, x) + b$

$A = sigmoid(z)$

$dz = A - Y$

$dw = \frac{1}{m} X \cdot dz^T$

$db = \frac{1}{m} \cdot np.sum(dz)$

$w = w - \alpha dw$

$b = b - \alpha db$

} 梯度下降的一次迭代

注: python的广播:

又按矩阵每列求和: $A.sum(axis=0)$ 竖直求和

$A.sum(axis=1)$ 水平求和

$x\_norm = np.linalg.norm(x, axis=1, keepdims=True)$ ⟶ $\|x\|$ (又对行)