Machine Learning

Logistic Regression

1. Model

logistic function:

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

hypothesis:

$$h_{ heta}(x) = g(heta^T x) = rac{1}{1 + e^{- heta^T x}}$$

maximum likelihood:

$$egin{split} P(y = 1 | x; heta) &= h_{ heta}(x) \ P(y = 0 | x; heta) &= 1 - h_{ heta}(x) \ P(y | x; heta) &= (h_{ heta}(x))^y (1 - h_{ heta}(x))^{1-y} \end{split}$$

input:

$$(x^{(1)},y^{(1)}),(x^{(2)},y^{(2)}),\cdots,(x^{(m)},y^{(m)}) ext{ where } y^{(i)} \in \{0,1\}$$

cost function:

$$L(heta) = \prod_{i=1}^m (h_ heta(x^{(i)}))^{y^{(i)}} (1 - h_ heta(x^{(i)}))^{1 - y^{(i)}} \ \ell(heta) = \log L(heta) = \sum_{i=1}^m y^{(i)} \log h_ heta(x^{(i)}) + (1 - y^{(i)}) \log (1 - h_ heta(x^{(i)}))$$

derivative:

$$egin{aligned} rac{\partial}{\partial heta_j} \ell(heta) &= \sum_{i=1}^m \left(y^{(i)} rac{1}{h_{ heta}(x^{(i)})} - (1-y^{(i)}) rac{1}{1-h_{ heta}(x^{(i)})}
ight) rac{\partial}{\partial heta_j} h_{ heta}(x^{(i)}) \ &= \sum_{i=1}^m \left(y^{(i)} (1-h_{ heta}(x^{(i)})) - (1-y^{(i)}) h_{ heta}(x^{(i)})
ight) x_j^{(i)} \ &= \sum_{i=1}^m (y^{(i)} - h_{ heta}(x^{(i)})) x_j^{(i)} \end{aligned}$$

gradient ascent:

$$egin{aligned} heta_j := heta_j + lpha rac{\partial}{\partial heta_j} \ell(heta) \ &= heta_j + lpha \sum_{i=1}^m (y^{(i)} - h_ heta(x^{(i)})) x_j^{(i)} \end{aligned}$$

stochastic gradient ascent:

$$heta_j := heta_j + lpha(y^{(i)} - h_ heta(x^{(i)}))x_j^{(i)}$$

Newton's method:

$$\theta := \theta - H^{-1} \nabla_{\theta} \ell(\theta)$$

where H is Hessian matrix, and $H_{ij}=rac{\partial^2\ell(heta)}{\partial heta_i\partial heta_i}$.

2. Softmax Regression

input:

$$(x^{(1)},y^{(1)}),(x^{(2)},y^{(2)}),\cdots,(x^{(m)},y^{(m)}) ext{ where } y^{(i)} \in \{1,\dots,K\}$$

hypothesis:

$$h_{ heta}(x) = egin{bmatrix} P(y=1|x; heta) \ P(y=2|x; heta) \ dots \ P(y=K|x; heta) \end{bmatrix} = rac{1}{\sum_{j=1}^K \exp(heta^{(j) op}x)} egin{bmatrix} \exp(heta^{(1) op}x) \ \exp(heta^{(2) op}x) \ dots \ \exp(heta^{(K) op}x) \end{bmatrix}$$

cost function:

$$L(heta) = \prod_{i=1}^m \sum_{k=1}^K 1\left\{y^{(i)} = k
ight\} rac{\exp(heta^{(k) op}x^{(i)})}{\sum_{j=1}^K \exp(heta^{(j) op}x^{(i)})} \ \ell(heta) = \log L(heta) = \sum_{i=1}^m \sum_{k=1}^K 1\left\{y^{(i)} = k
ight\} \log rac{\exp(heta^{(k) op}x^{(i)})}{\sum_{j=1}^K \exp(heta^{(j) op}x^{(i)})}$$

derivative:

$$abla_{ heta^{(k)}}\ell(heta) = \sum_{i=1}^m \left[x^{(i)} \left(1\{y^{(i)} = k\} - P(y^{(i)} = k|x^{(i)}; heta)
ight)
ight]$$

gradient ascent:

$$egin{aligned} heta^{(k)} &:= heta^{(k)} + lpha
abla_{ heta^{(k)}} \ell(heta) \ &= heta^{(k)} + lpha \sum_{i=1}^m \left[x^{(i)} \left(1\{y^{(i)} = k\} - P(y^{(i)} = k|x^{(i)}; heta)
ight)
ight] \end{aligned}$$

stochastic gradient ascent:

$$heta^{(k)} := heta^{(k)} + lpha \left[x^{(i)} \left(1\{y^{(i)} = k\} - P(y^{(i)} = k | x^{(i)}; heta)
ight)
ight]$$