Logistic Regression 2

Goals: Understand logistic loss better, compare LS
Know gradient algorithm

Recap: Probability model logistic regression

$$-\log P[\vec{y} \mid X, \vec{\beta}]$$

$$= \sum_{i=1}^{n} \log\left(1 + e^{-y_i \vec{x}_i^T \vec{\beta}}\right)$$

<u>Defn</u> The <u>logistic loss</u> $L_{logistic}(z) = \log(1+e^{-z})$

<u>Defn</u> We call $z_i = y_i f(\vec{x}_i)$ the <u>margin</u> of example i.

$$= y_i \vec{x}_i^T \vec{\beta} \qquad (y_i = \pm 1)$$

A loss function measures error between predictions ($f(\vec{x})$) and truth (labels y).

- usually result from log-likelihood

We've seen another loss function, least squares

$$L_{LS}(y, \underbrace{f(\vec{x})}_{\hat{y}}) = (y - f(\vec{x}))^2$$

$$= (1 - \underbrace{y f(\vec{x})}_{z})^2 \qquad \text{If } y = \pm 1$$

$$= (1 - z)^2 = L_{LS}(z)$$

Margin: $z = y\, f(\vec{x})$

$f(\vec{x}) > 0$ predict $\hat{y} = +1$

$f(\vec{x}) < 0$ predict $\hat{y} = -1$

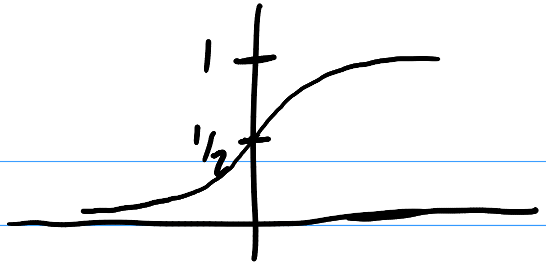
positive margin = good, $\hat{y} = y$

negative margin = bad, $\hat{y} \neq y$

$$L_{01}(\hat{y}, y) = \mathbb{1}_{\{\hat{y} \neq y\}} = \mathbb{1}_{\{\operatorname{sgn}(f(\vec{x})) = y\}}$$

Captures the confidence effect on errors

Recipe: Have a loss function
Minimize loss over $\vec{\beta}$



$$\vec{\beta}_{\text{logistic}} = \arg\min_{\vec{\beta}} L_{\text{logistic}}$$

$$= \arg\min_{\vec{\beta}} \sum_{i=1}^{n} \log\left(1 + e^{-y_i \vec{x}_i^T \vec{\beta}}\right)$$

$$\nabla_{\vec{\beta}} L_{\text{logistic}} = -\sum_{i=1}^{n} \overbrace{P[Y_i = -y_i \mid X, \vec{\beta}]}^{\text{Probability of making a mistake}} \, y_i \vec{X}_i$$

$$\nabla_{\vec{\beta}} L_{LS} = 2 \sum_{i=1}^{n} \left(\vec{x}_i^T \vec{\beta} - y_i\right) \vec{x}_i$$

points in direction of $\vec{x}_i$

Assuming $n=1$

$f=-4 \quad y=-1$

$\nabla L_{logistic} \simeq 0$

ex/ $f(\vec{x})=-4, \quad y=+1$

predict $Y=-1$

$\nabla_\beta L_{logistic} \simeq -1 \cdot \vec{x}$

only cares about mistakes

$\nabla_\beta L_{LS} = 2(-4-1)\vec{x} = -10\vec{x}$

ex/ $f(\vec{x})=+4 \quad y=+1$

predict $Y=+1$

really doesn't care about large margin

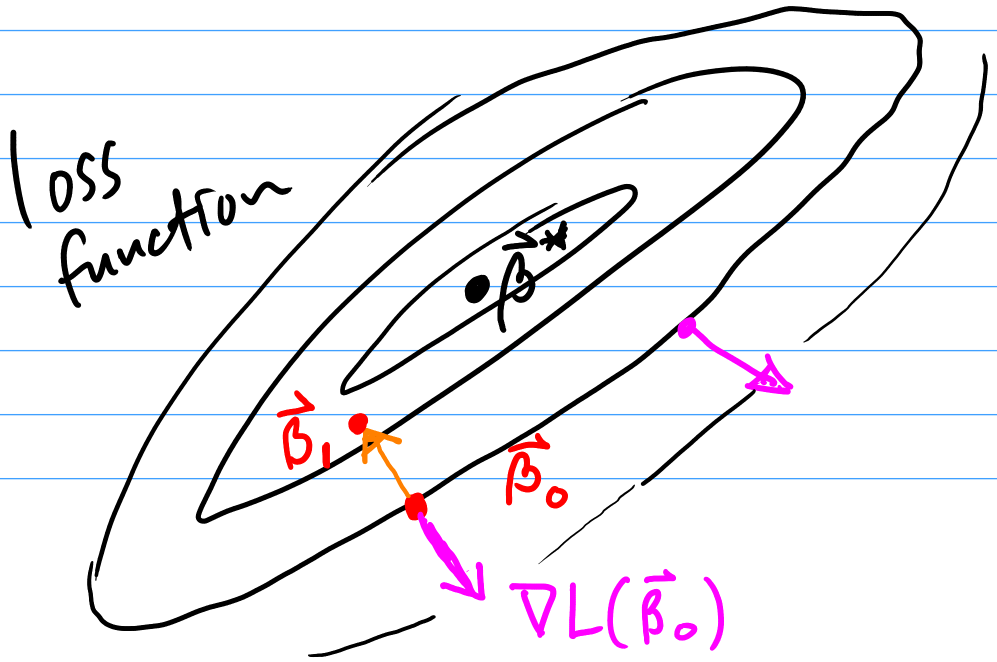
$\nabla_\beta L_{logistic} \simeq 0 \cdot \vec{x}$

$\nabla_\beta L_{LS} = 2(4-1)\vec{x} \simeq 6\vec{x}$

bad effect on large margin

Because of nonlinearity in logistic function, cannot set $\nabla_{\vec{\beta}} L_{logistic} = 0$ and solve in one step.

Need iterative method to find minimizer



gradient point in steepest direction of increase of L

$-$gradient points in steepest descent direction

Gradient descent

Initialize: $\vec{\beta}_0$ initial guess

h step size

L loss function, ∇L

for $t = 1, 2, \ldots$

$$\vec{\beta}_t = \vec{\beta}_{t-1} - h \nabla_{\vec{\beta}} L(\vec{\beta}_{t-1})$$

if converge

break

Return $\vec{\beta}_t$

See the interactive webpage on Piazza