Recipe: fit parameters to data

Loss and function: $L(y^{(n)}, f(\underline{x}^{(n)}, \underline{w})) = L_n$

Square loss: $(y^{(n)} - f_n)^2$

Negative log probability: $-\log p(y^{(n)} \mid \underline{x}^{(n)}, \underline{w})$

Cost

$$C = \sum_n L_n \quad \text{sum over training set}$$

Learning direction

$$-\frac{1}{N} \nabla_{\underline{w}} C = \underbrace{-\frac{1}{N} \sum_n \nabla_{\underline{w}} L_n}$$

Monte Carlo approx:

$-\nabla_{\underline{w}} L_n$ for random n

$\Rightarrow$ Identify probabilistic model of y, $p(y \mid \underline{x}, \underline{w})$

$\Rightarrow$ Update parameters:

$$\underline{w} \leftarrow \underline{w} + \eta \nabla_{\underline{w}} \log p(y^{(n)} \mid \underline{x}^{(n)}, \underline{w})$$

S.G.D. on loss. S.G. Ascent on log likelihood

Need Likelihood of $\underline{w}$, ϵ

$$P(y=1 \mid \underline{x}, \underline{w}, \epsilon) = \sum_{m \in \{0,1\}} P(y=1, m \mid \underline{x}, \underline{w}, \epsilon)$$

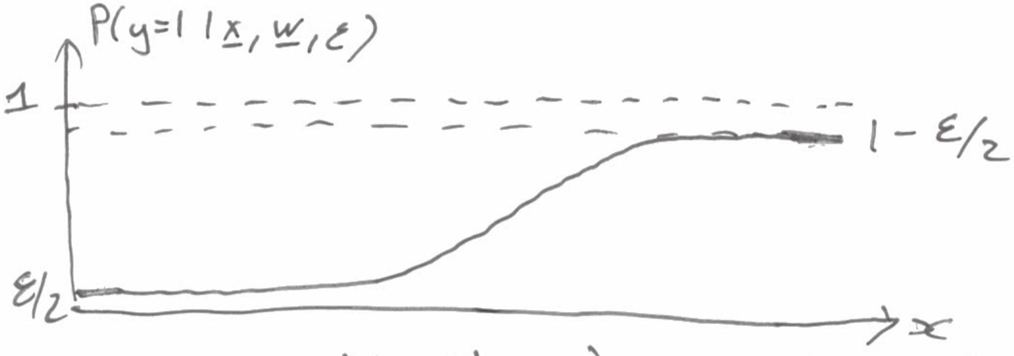
(Sum Rule)

$$= \sum_{m \in \{0,1\}} P(y=1 \mid \underline{x}, \underline{w}, \not{\epsilon}, m)\, P(m \mid \not{\underline{x}}, \not{\underline{w}}, \epsilon)$$

For this model

(Product Rule)

$$= \underbrace{(1-\epsilon)\,\sigma(\underline{w}^{\top}\underline{x})}_{m=1} + \underbrace{\epsilon/2}_{m=0}$$

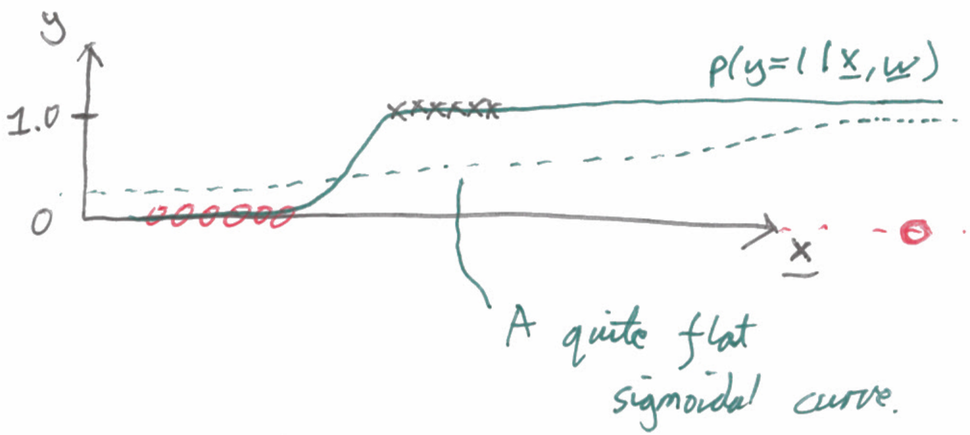


$$\nabla_{\underline{w}} \log P(y^{(n)} \mid \underline{x}^{(n)}, \underline{w}, \epsilon) = \ldots \text{calculus/algebra}$$

$$= \frac{1}{1 + \frac{1}{2}\left(\frac{\epsilon}{1-\epsilon}\right)\frac{1}{\sigma_n}} \nabla_{\underline{w}} \log \underbrace{\sigma_n}_{(p(y \mid \underline{w}, \underline{x})}$$

Logistic regression

Robust Logistic Regression



Each example has binary variable

$$m^{(n)} \in \{0, 1\}$$ hidden variable / latent

I will assume:

$$p(m|\varepsilon) = \begin{cases} 1-\varepsilon & m=1 \\ \varepsilon & m=0 \end{cases}$$

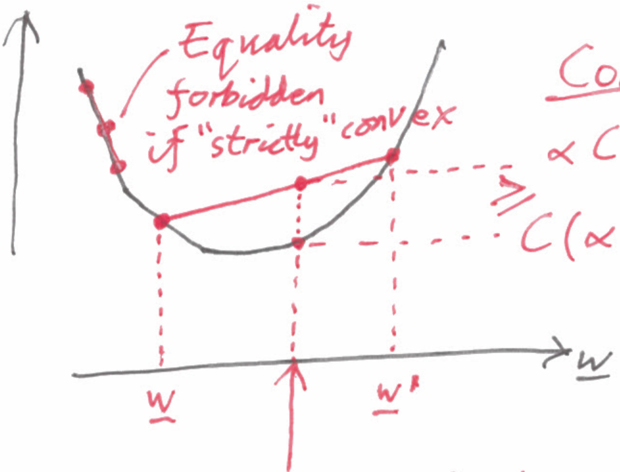
$\varepsilon = 0.01$

Model for labels:

$$P(y=1|\underline{x}, \underline{w}, m) = \begin{cases} \sigma(\underline{w}^T \underline{x}) & m=1 \\ 1/2 & m=0 \end{cases}$$

Convex Optimization problem?

Cost $C(\underline{w})$



Convex:

$\alpha C(\underline{w}) + (1-\alpha) C(\underline{w}')$

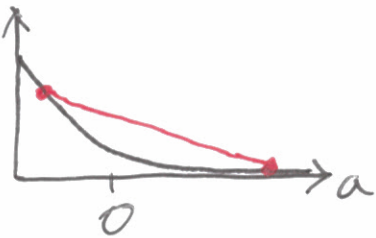
$\geq C(\alpha \underline{w} + (1-\alpha)\underline{w}')$

for all $\underline{w}, \underline{w}', \alpha$

$\alpha \underline{w} + (1-\alpha)\underline{w}', \quad 0 \leq \alpha \leq 1$

Logistic Regression

$-\log \sigma(a)$



Cost is sum
convex functions
$\Rightarrow$ convex

$-N(a; 0, \sigma^2)$

Not convex

Robust logistic regression

Not convex
$(\epsilon > 0)$

→ Fit anyway?

How do we fit ε?

By hand? Grid of settings $\varepsilon \in [0,1]$

maybe on a log scale 0.1, 0.01, 0.001

Jointly fit $\underline{\theta} = \begin{bmatrix} \underline{w} \\ \varepsilon \end{bmatrix}$, $\nabla_{\underline{\theta}} C$

Not convex, try anyway.

Doesn't work.

$$\varepsilon \in [0,1]$$

Trick: reparameterize model

$$\varepsilon = \sigma(b)$$

$\uparrow$ $-\infty < b < \infty$

derive $\frac{\partial C}{\partial b}$ and optimize $\begin{bmatrix} \underline{w} \\ b \end{bmatrix}$