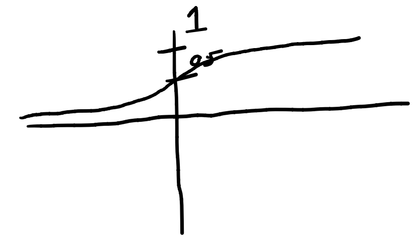
Logistic Regression

Discriminative probabilistic model $P(y|\bar{x})$

(generative: $P(\bar{x}, y)$)

logistic

$$\frac{e^x}{1+e^x}$$



$$P(y=+1|\bar{x}) = \frac{e^{\bar{w}^T f(\bar{x})}}{1+e^{\bar{w}^T f(\bar{x})}}$$

$$P(y=-1|\bar{x}) = \frac{1}{1+e^{\bar{w}^T f(\bar{x})}}$$

Decision boundary?

return $+1$ if $P(y=+1|\bar{x}) > 0.5$

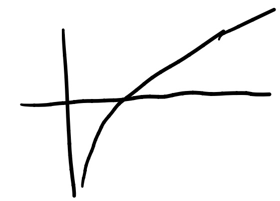
$\bar{w}^T f(\bar{x}) > 0$

Training

For a dataset $\left(\bar{x}^{(i)}, y^{(i)}\right)_{i=1}^{D}$, want to maximize

$$\prod_{i=1}^{D} P\left(y^{(i)} \mid \bar{x}^{(i)}\right)$$ maximum likelihood

$$\max_{\bar{w}} \sum_{i=1}^{D} \log P\left(y^{(i)} \mid \bar{x}^{(i)}\right)$$ log likelihood



$$\Rightarrow \min_{\bar{w}} \sum_{i=1}^{D} \underbrace{-\log P\left(y^{(i)} \mid \bar{x}^{(i)}\right)}_{\text{loss}\left(\bar{x}^{(i)}, y^{(i)}, \bar{w}\right)}$$

training objective
negative log likelihood
(NLL)

Need to compute $\frac{\partial}{\partial \bar{w}} \text{loss}\left(\bar{x}^{(i)}, y^{(i)}, \bar{w}\right)$

LR Gradients

$\boxed{\text{Assume } y^{(i)} = +1}$

$$\frac{\partial}{\partial \overline{w}} \text{loss} = \frac{\partial}{\partial \overline{w}} \left[-\overline{w}^T f(\overline{x}) + \log\left(1 + e^{\overline{w}^T f(\overline{x})}\right) \right]$$

$$= -f(\overline{x}) + \frac{1}{1 + e^{\overline{w}^T f(\overline{x})}} \cdot e^{\overline{w}^T f(\overline{x})} \cdot f(\overline{x})$$

$$= f(\overline{x}) \left[-1 + \frac{e^{\overline{w}^T f(\overline{x})}}{1 + e^{\overline{w}^T f(\overline{x})}} \right] = f(\overline{x}) \left[P(y=+1 \mid \overline{x}) - 1 \right]$$

Update: $\overline{w} \leftarrow \overline{w} + \alpha f(\overline{x}) \left(1 - P(y=+1 \mid \overline{x}) \right)$

$P(y=+1 \mid \overline{x}) \approx 1$: no update

$P(y=+1 \mid \overline{x}) \approx 0$: perceptron update

Let $z = \overline{w}^T f(\overline{x})$

NLL : $\log(1 + e^{z}) - z$