Logistic Regression

Model $y \mid \underset{\sim}{x}$ directly

$\hat{f}(\underset{\sim}{x}) =$
$\hat{y} = \underset{c}{\operatorname{argmin}}\, \hat{p}(y \mid \underset{\sim}{x})$

Basic version: binary classifier $y \in \{0, 1\}$

Prediction fn:

Since $y \in \{0,1\}$

$$p(y=1 \mid \underset{\sim}{x}) = 1 - p(y=0 \mid \underset{\sim}{x})$$

only need to model one

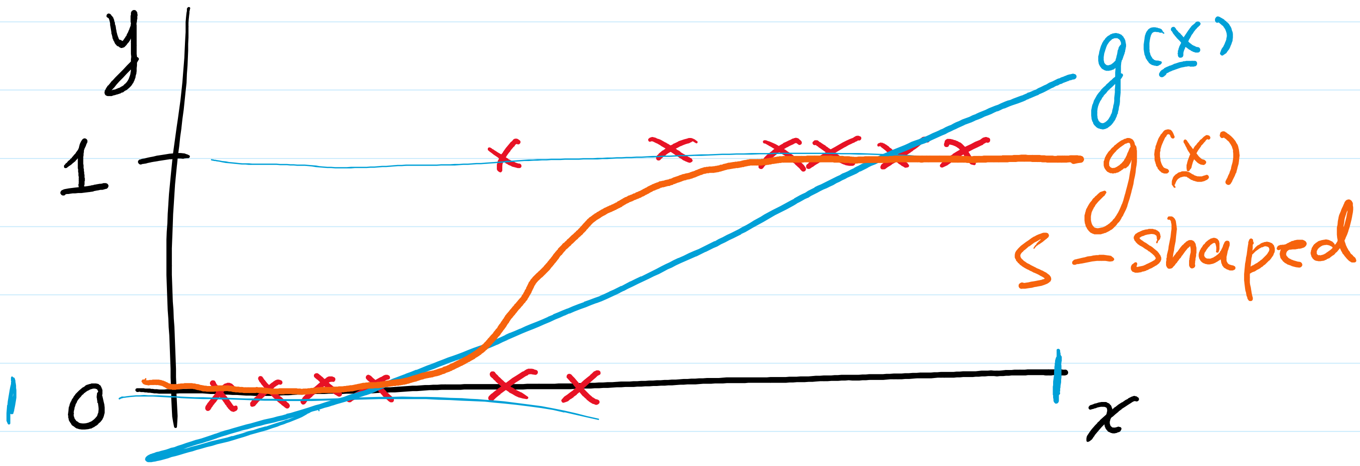
Equiv: $\hat{f}(\underset{\sim}{x}) = 1 \iff \boxed{p(y=1 \mid \underset{\sim}{x})} > p(y=0 \mid \underset{\sim}{x})$

$$> 1 - p(y=1 \mid \underset{\sim}{x})$$

$$\iff p(y=1 \mid \underset{\sim}{x}) > 1/2.$$

$$Y \mid \underset{\sim}{X} = \underset{\sim}{x} \sim \text{Bern}(\ ???\ )$$

$$p(y=1 \mid \underset{\sim}{x}) = g(\underset{\sim}{x})$$



In particular logistic regression uses logistic curve:

$$g(\underset{\sim}{x}) = \text{logistic}(\underset{\sim}{x}^T \beta)$$

$$\boxed{\text{logistic}(x) = \frac{1}{1+e^{-x}}}$$

<u>inverse</u>: $\text{logit}(a) = \log\left(a/(1-a)\right)$

Notice: $\delta_1(\underset{\sim}{x}) = p(y=1|\underset{\sim}{x}) = g(\underset{\sim}{x}) = \text{logistic}(\underset{\sim}{x}^T\beta)$

then $\text{logit}(\delta_1(\underset{\sim}{x})) = \underset{\sim}{x}^T\beta$

increasing

So logistic regression is a linear classifier.

How do I learn good values of β? MLE

$$\hat{\beta} = \underset{\beta}{\text{argmax}}\, \ell(\beta) \doteq \underset{\beta}{\text{argmin}}\, NLL(\beta)$$

$$\boxed{y_n | \underset{\sim}{x}_n \sim \text{Bern}(g(\underset{\sim}{x}_n))}$$

$$g(\underset{\sim}{x}_n) = \text{logistic}(x_n^T\beta)$$

$z \sim \text{Bern}(p)$, $\quad p(t) = \underline{p^t(1-p)^{1-t}}$

$$p = p(z=1)$$

$$L(\beta) = \prod_n p(\underset{\uparrow}{y_n} \mid \underset{\sim}{x}_n) \quad (Y_n = y_n)$$

$$= \prod_n g(\underset{\sim}{x}_n)^{y_n} (1 - g(\underset{\sim}{x}_n))^{1-y_n}$$

$$\ell(\beta) = \sum_n \left[y_n \log(g_\beta(\underset{\sim}{x}_n)) + (1-y_n) \log(1 - g_\beta(\underset{\sim}{x}_n)) \right]$$

$$= -\sum_n H(y_n, g(\underset{\sim}{x}_n))$$

cross entropy

$$H(p, q) = p \log(q) + (1-p) \log(1-q)$$

$$NLL(\beta) = \sum_n \boxed{H(y_n, g_\beta(\underset{\sim}{x}_n))}$$

cross entropy loss

$$\hat{\beta} = \underset{\beta}{\arg\min}$$

= minimize cross entropy

No closed form soln, use numerical optim.

<u>Gradient descent</u>:

Initialize guess $\hat{\beta}^{(0)}$
For $t = 1, 2, 3, \ldots$

$$\hat{\beta}^{(t)} \leftarrow \hat{\beta}^{(t-1)} - \alpha \nabla_{\beta} NLL(\hat{\beta}^{(t-1)})$$

Can extend to more than 1 class

Called <u>Multinomial Logistic Regression</u>

LDA v. Logistic Regression

LDA	Logistic Regression
1 model $\underline{X} \mid Y$ and Y normality assump. on $\underline{X}$	1 model $Y \mid \underline{X}$ directly no assump. about dist of $\underline{X}$
2 easier to fit, less issues	2 harder to fit, could have problems if classes very separated

3 both linear