Logistic Regression — classification

the "least-squares" of

Linear predictors: hyperplane fit

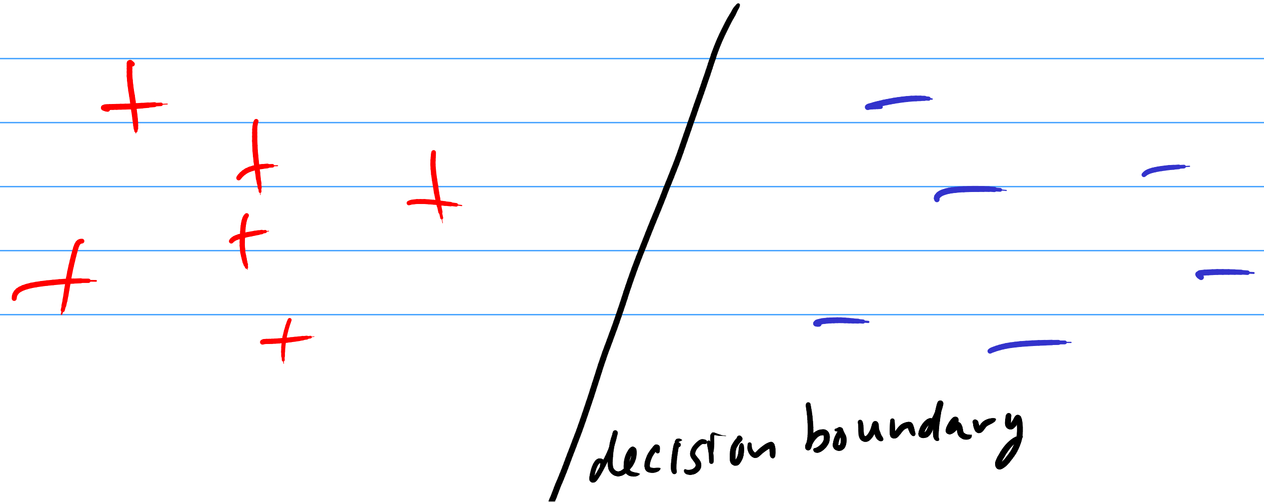
in regression: $f(\vec{x}) = \vec{x}^T \vec{\beta} \approx y$

"sign function"

in classification:

$$f(\vec{x}) = \vec{x}^T \vec{\beta}, \quad y = \operatorname{sgn}(f(\vec{x})) = \begin{cases} +1 & \text{if } f > 0 \\ -1 & \text{if } f < 0 \end{cases}$$

decision boundary



Connection to probability $P[Y=+1 \mid \vec{x}]$

log-odds : logit function

$$\log\left(\frac{P[Y=+1 \mid \vec{x}]}{P[Y=-1 \mid \vec{x}]}\right) = f(\vec{x})$$

ex/ $\log(x) = 0$ if $x = 1$

since $e^{\log x} = x = e^{0} = \underline{1}$

$$x = 1 \iff P[Y=+1 \mid \vec{x}] = P[Y=-1 \mid \vec{x}]$$

This means $f(\vec{x}) = 0$ is the decision boundary ex/ f $\overset{\text{small}}{\text{huge}}$, say $\overset{-100}{100}$

$$\cancel{\log}\left(\frac{P[Y=+1 \mid \vec{x}]}{P[Y=-1 \mid \vec{x}]}\right) = \underset{e}{\overset{e^{-100}}{100}}$$

$$P[Y=+1 \mid \vec{x}] = p$$

$$P[Y=-1 \mid \vec{x}] = 1-p$$

have to add up to 1

logit function (of p)

$$\log\left(\frac{p}{1-p}\right) = f$$

$e^{(\cdot)}$

$$\frac{p}{1-p} = e^{f}$$

$(1-p)$

$$p = (1-p)\, e^{f}$$

$+\, p e^{f}$

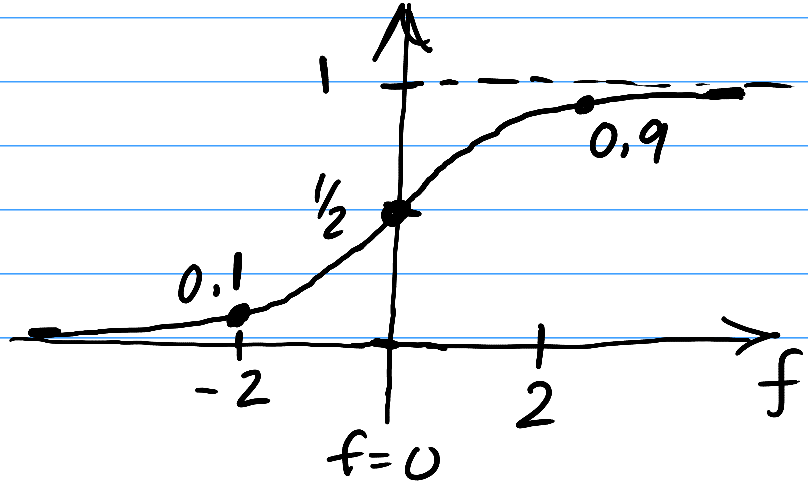
$$p + p \cdot e^{f} = e^{f}$$

$$p\left(1+e^{f}\right) = e^{f}$$

logistic function (of f)

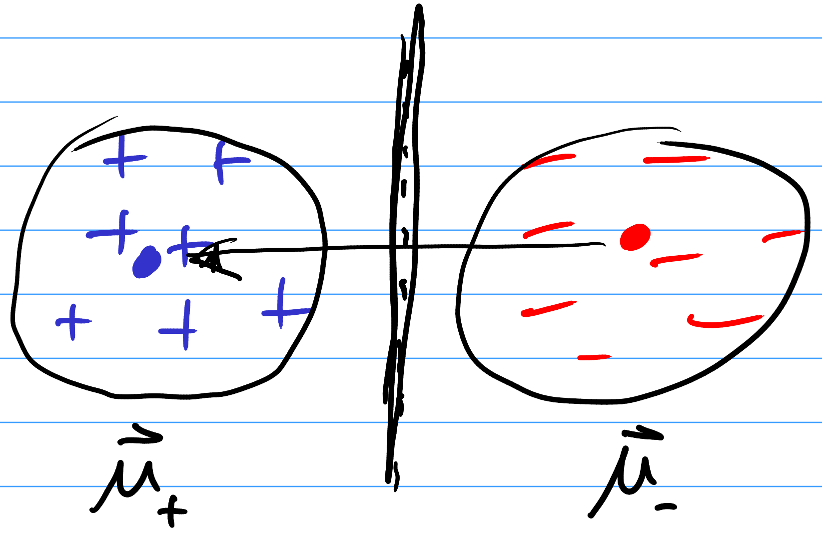
$$p = \frac{1}{1+e^{-f}}$$

$$= P[\, Y=+1 \,/\, \vec{x} \,]$$



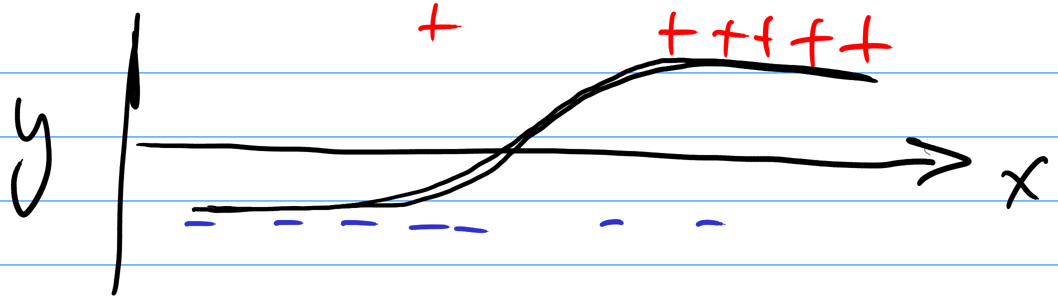
Connection to Gaussian: LDA

linear discriminant analysis



equivalent to logistic

$$(\vec{\mu}_+ - \vec{\mu}_-)\, c = \vec{\beta}$$



Probabilistic interpretation of least squares:
equivalent to maximizing likelihood under Gaussian assumptions

Lets do same thing. Likelihood of data under logistic model. Data $(\vec{x}_i, y_i)_{i=1}^{n}$, model $f(\vec{x}) = \vec{x}^T\vec{\beta}$

$$\underbrace{P[\underset{\text{all } y_i\text{'s}}{\vec{y}} \mid \vec{\beta}, X]}_{\text{likelihood of } \vec{\beta}} = \prod_{i=1}^{n} P[Y = y_i \mid \vec{\beta}, \vec{x}_i]$$

random variable that's the label

realization, what you measure

Just like before max likelihood $\Longleftrightarrow$ min (- log likelihood)

$$\left(\log(a \cdot b) = \log a + \log b \right)$$

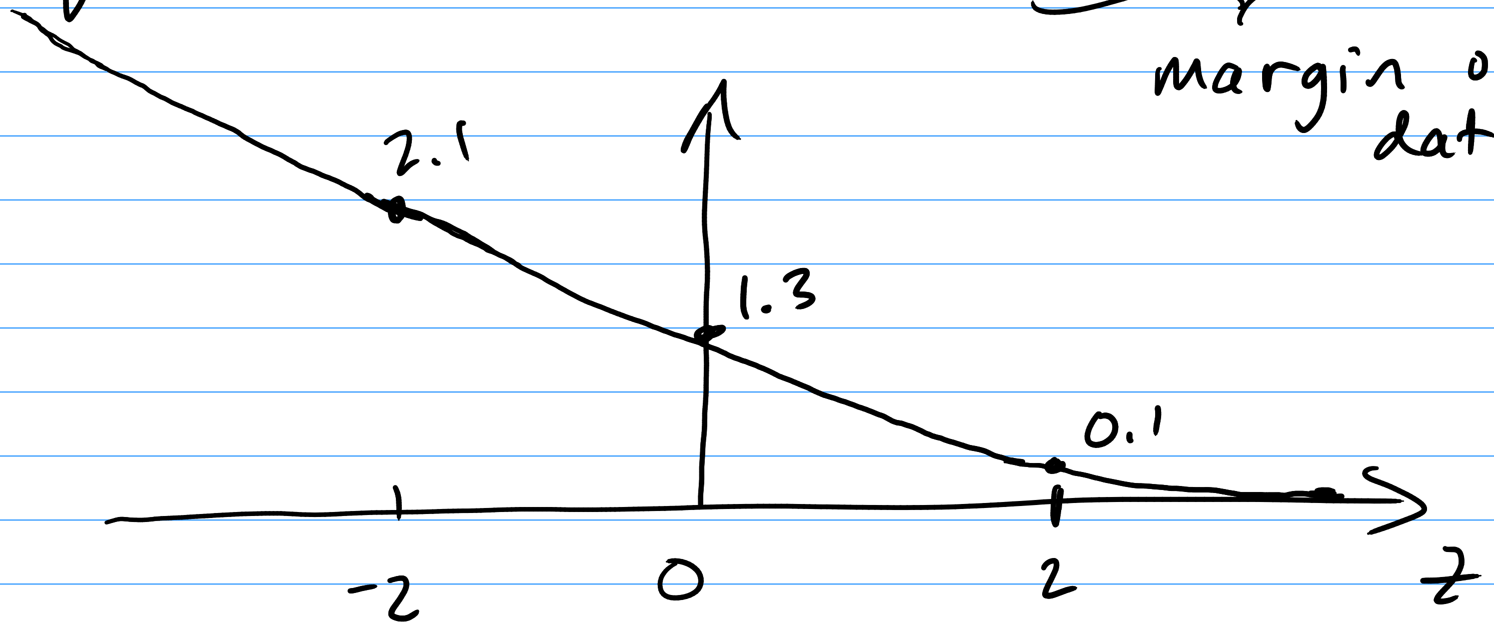
$$-\log P[\vec{y} \mid X, \vec{\beta}] = -\sum_{i=1}^{n} \log P[Y = y_i \mid \vec{x}_i, \vec{\beta}]$$

logistic formula

$$= -\sum_{i=1}^{n} \log \begin{cases} \dfrac{1}{1+e^{-\vec{x}_i^T \vec{\beta}}} & \text{if } y_i = +1 \\ \dfrac{1}{1+e^{\vec{x}_i^T \vec{\beta}}} & \text{if } y = -1 \end{cases}$$

$$= \sum_{i=1}^{n} \underbrace{\log\left(1 + \exp(-y_i \vec{x}_i^T \vec{\beta})\right)}_{\text{logistic loss function}}$$

$$\log\left(1+e^{-z}\right), \quad z = \underbrace{y_i f(\vec{x}_i)}_{\text{margin of } i^{th} \text{ data pt.}}$$



ex/ $z=0$, $\log(1+e) \approx 1.3$

$z \to \infty$, ≈ 0

$z = 2$, ≈ 0.1

$z = -2$, ≈ 2.1

Next step

$$\min_{\vec{\beta}} \underbrace{(-\text{log-likelihood})}_{\text{loss function}}$$

$$\text{loss}(\vec{\beta}\,;\,X, \vec{y})$$

maximum likelihood estimation

ex/ $\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} x_i$

ex/ OLS