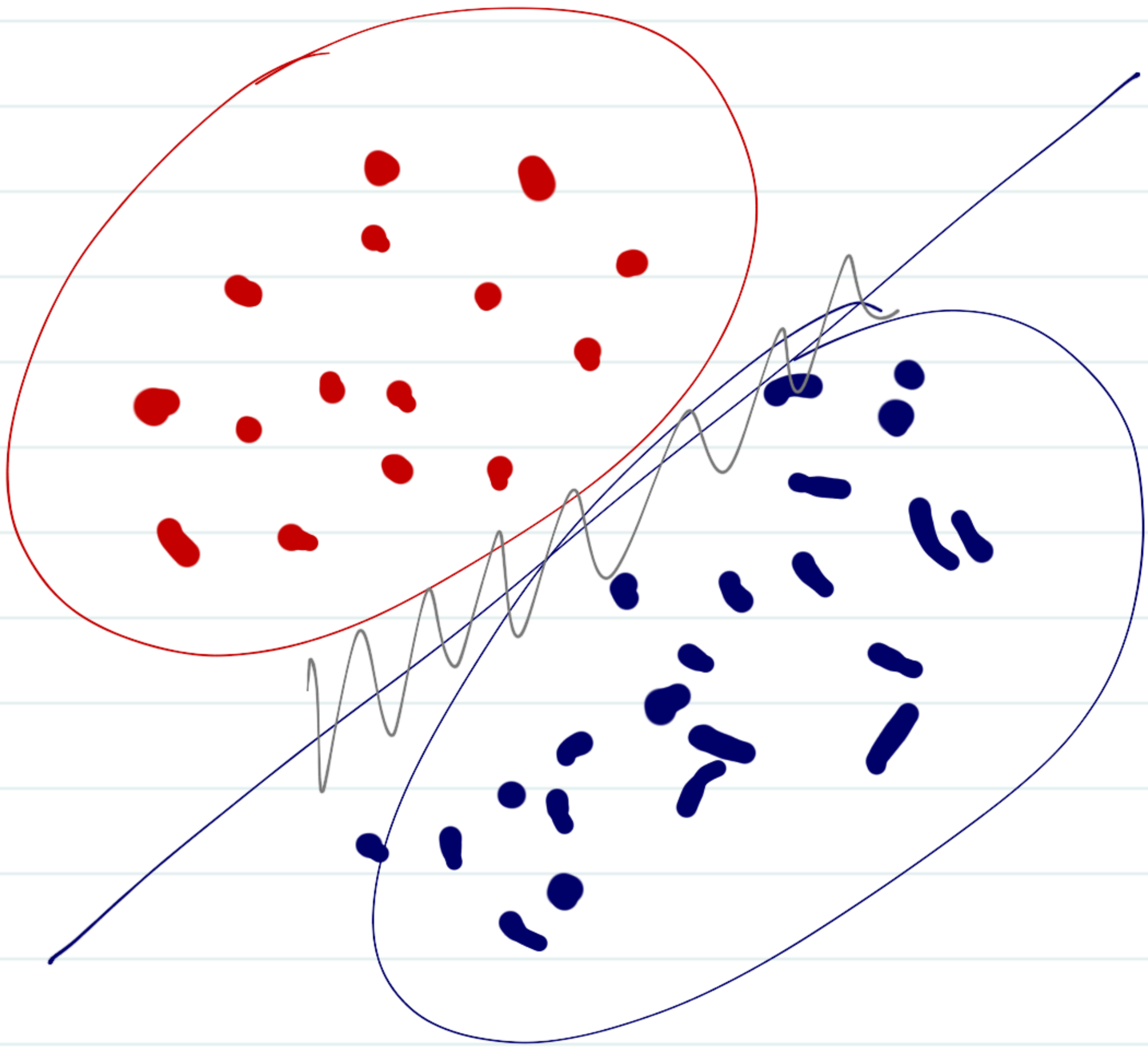
Regularized logistic regression

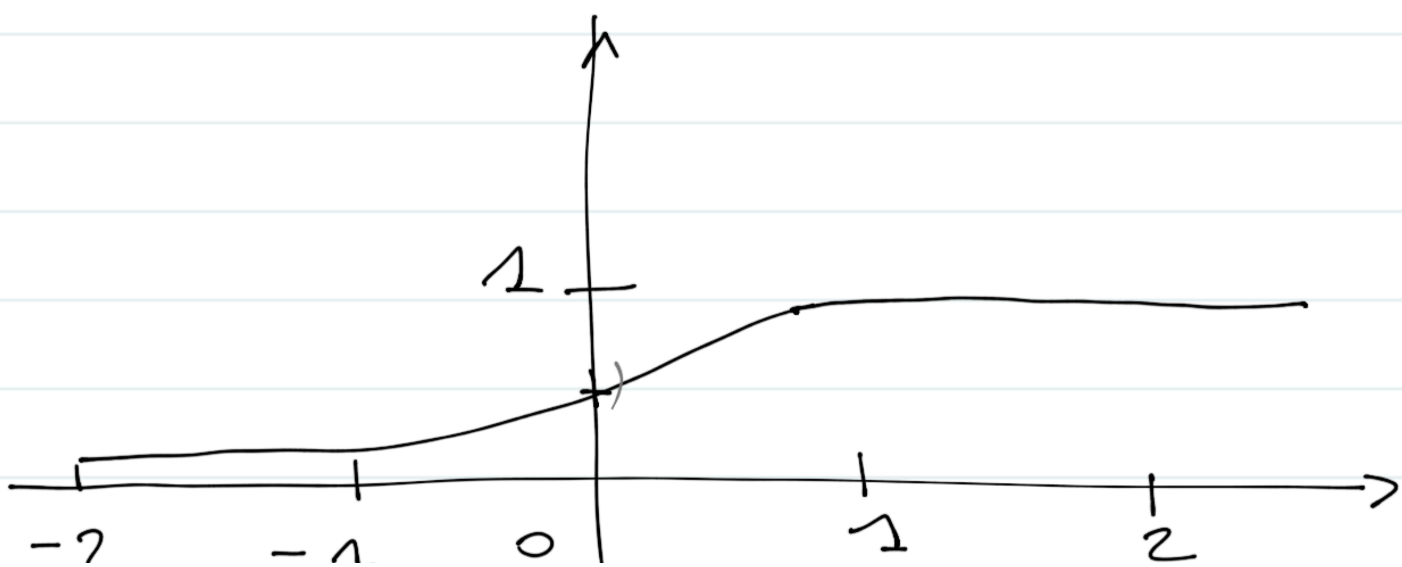
$$\min_{\beta \in \mathbb{R}^d} \frac{1}{n} \sum_{i=1}^{n} \log\left(1 + e^{-y_i x_i^T \beta}\right) + \lambda \|\beta\|_2^2$$



$$\boxed{P(Y = 1 \mid x ; \beta) = g(x^T \beta)}$$

$$g(z) = \frac{1}{1 + e^{[redacted]z}}$$

- logistic function
- sigmoid function



Either you are in class $+1$

or you are in class -1

$$P(Y=+1 \mid x;\beta) + P(Y=-1 \mid x;\beta) = 1$$

$$P(Y=-1 \mid x;\beta)$$

$$= 1 - P(Y=+1 \mid x;\beta)$$

$$\log\left(\frac{P(Y=+1 \mid x;\beta)}{P(Y=-1 \mid x;\beta)}\right)$$

$$= x^T\beta$$

$$1 - g(z) = g(-z)$$

$$g(-z) = \frac{1}{1 + e^{-(-z)}}$$

$$= \frac{1}{1 + e^{+z}}$$

$$= \frac{e^{-z} \cdot 1}{e^{-z}(1 + e^{+z})}$$

$$= \frac{e^{-z}}{e^{-z} + e^{-z}e^{+z}}$$

$$= \frac{e^{-z}}{1 + e^{-z}}$$

$$g(-z) = \frac{e^{-z}}{1+e^{-z}}$$

I want to prove that

$$1 - \boxed{g(z)} = g(-z)$$

$$g(z) = \frac{1}{1+e^{-z}}$$

$$g(-z) = \cancel{\frac{1+e^{-z}}{1+e^{-z}}} - \frac{1}{1+e^{-z}}$$

$$= 1 - g(z)$$

$\square$

$$P(Y=-1 \mid x; \beta)$$

$$= 1 - P(Y=+1 \mid x; \beta)$$

$$= 1 - g(+x^T \beta) \leftarrow \boxed{\text{PROPERTY}}$$

$$= g(-x^T \beta)$$

$P(Y=+1 \mid x; \beta) = g(+1 \cdot x^T \beta$

$P(Y=-1 \mid x; \beta) = g(-1 \cdot x^T \beta$

so

$P(y \mid x; \beta) = g(y \cdot x^T \beta)$

Likelihood $\mathcal{L}$

$$\mathcal{L}((x_1, y_1), \ldots, (x_n, y_n); \beta)$$

$$= \prod_{i=1}^{n} P(y_i \mid x_i; \beta)$$

$$\max_{\beta} \mathcal{L}(\beta)$$

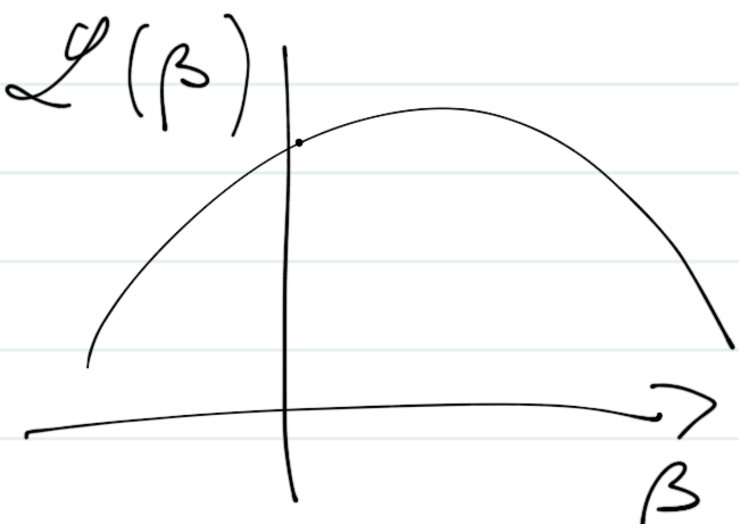
$$\min_{\beta} -\log(\mathcal{L}(\beta$$

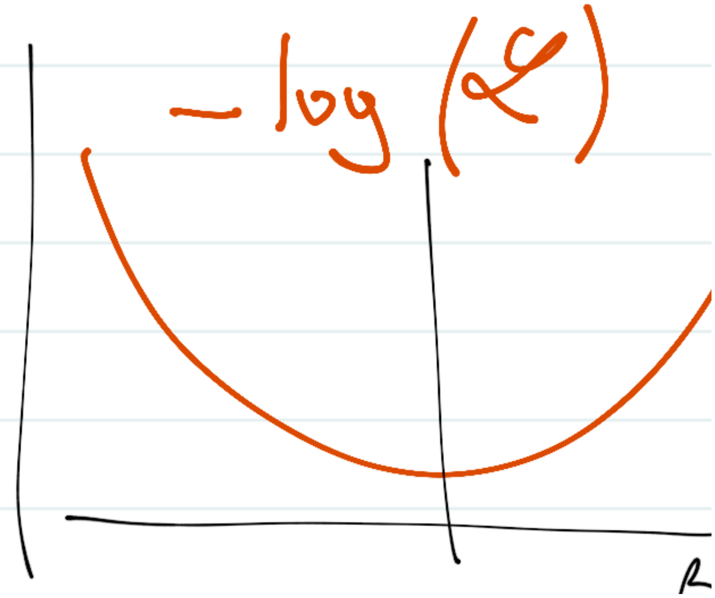
$$\log(a \times b) = \log(a) + \log(b)$$

$$\log\left(\prod_{i=1}^{n} p(y_i \mid x_i ; \beta)\right)$$

$$= \log\left(p(y_1 \mid x_1 ; \beta) \times \cdots \times p(y_n \mid x_n ; \beta)\right)$$

$$= \sum_{i=1}^{n} \log\left(p(y_i \mid x_i ; \beta)\right)$$





$$-\log(\mathcal{L}) = n \text{ Empirical Risk}$$

$$1 - g(z) = g(-z)$$

PROPERTY

$$-\log(\mathcal{L})$$

$$= \sum_{i=1}^{n} -\log\{p(y_i \mid x_i ; \beta)\}$$

$$= \sum_{i=1}^{n} -\log\{g(y_i \, x_i^T\beta)\}$$

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

Empirical Risk

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

- $y_i = +1$

$$g(y_i x_i^T \beta)$$

$$= g(x_i^T \beta)$$

- $y_i = -1$

$$g(-1 \cdot x_i^T \beta)$$

$$= 1 - g(x_i^T \beta)$$

RECALL THE PROPERTY

$$\frac{\partial}{\partial \beta} \log\left(1 + e^{\left(-y_i \cdot x_i' \beta\right)}\right)$$

$$\frac{\partial}{\partial \beta}\left(-y_i \cdot x_i^T \beta\right)$$

$$= -y_i \cdot x_i$$

$$f(x) = \overline{\exp(x)}$$

$$f'(x) = \exp(x)$$

$$\left(\log(u)\right)' = + \frac{u'}{u}$$

$$\log\left(1 + e^{-y x' \beta}\right)$$

- $f(\cdot) = \log(\cdot)$
- $g(\cdot) = 1 + e^{-\cdot}$
- $h(\cdot) = y \cdot x^T \beta$

$$\to f(g(h(\beta)))$$

$$= f((g \circ h)(\beta))$$

$$(f \circ g \circ h)' = (g \circ h)' \, f'((g \circ h))$$

$$(g \circ h)' = h' \cdot g'(h)$$

$$(f \circ g \circ h)'$$

$$= \underbrace{h'} \cdot g'(h) \cdot f'(g \circ h)$$

$$y \cdot x \cdot \left(-1 \cdot e^{-y \cdot x^T \beta}\right) \frac{1}{1 + e^{-y \cdot x^T \beta}}$$

$$\frac{\partial}{\partial \beta} \log\left(1 + e^{-y \cdot x^T \beta}\right)$$

$$= -y \cdot x \frac{e^{-y x^T \beta}}{1 + e^{-y \cdot x^T \beta}}$$

$$1 - \boxed{p_i} = \frac{e^{-y_i \cdot x_i^T \beta}}{1 + e^{-y_i \cdot x_i^T \beta}}$$

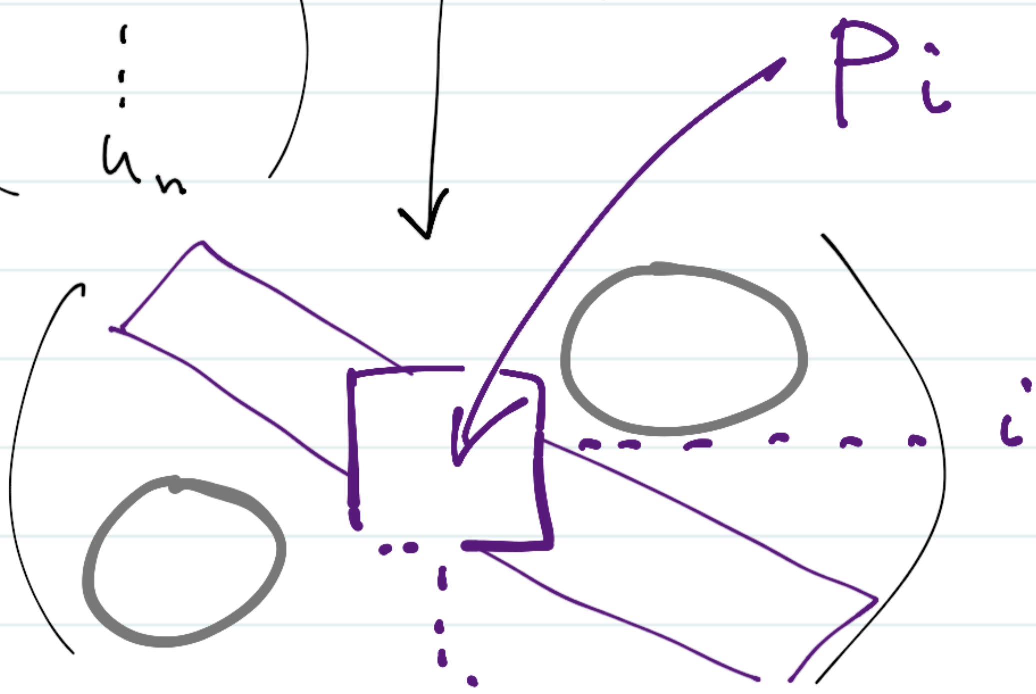
$$\frac{\partial}{\partial \beta} \log\left(1 + e^{-y_i \cdot x_i^T \beta}\right)$$

$$= -y_i \cdot x_i \cdot (1 - p_i)$$

INTERLUDE

$$u = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix} \quad n$$

D =



$$\mathrm{Diag}\left([p_1, \dots, p_n] \right)$$

$$= \begin{pmatrix} p_1 & & 0 \\ & \ddots & \\ 0 & & p_n \end{pmatrix} = D$$

$$Du = \sum_{i=1}^{n} p_i \, u_i$$

$$\frac{\partial}{\partial \beta} \left(R_{emp}(\beta) \right) = \frac{\partial}{\partial \beta} \left(\frac{1}{n} \sum_{i=1}^{n} \log \dots \right)$$

$$P = Id - Diag([p_1, \dots, p_n])$$

$$= \begin{pmatrix} 1-p_1 & & 0 \\ & \ddots & \\ 0 & & 1-p_n \end{pmatrix}$$

$$\frac{\partial}{\partial \beta}\left(R_{emp}(\beta)\right) = -\frac{1}{n} X \quad Y P$$

$$\frac{\partial}{\partial \beta}\left(\lambda \, \|\beta\|_2^2\right) = 2\lambda\beta$$

$$\boxed{\frac{\partial f}{\partial \beta} = -\frac{1}{n} X \quad YP \; + \; 2\lambda\beta}$$