Modelling game outcomes

Two player game → win / lose (no draws)

Player 1	Player 2	Winner
Fred	Nina	Nina
Jo	Tina	Jo
Tina	Nina	Nina
...	...	...
Fred	Nina	Fred

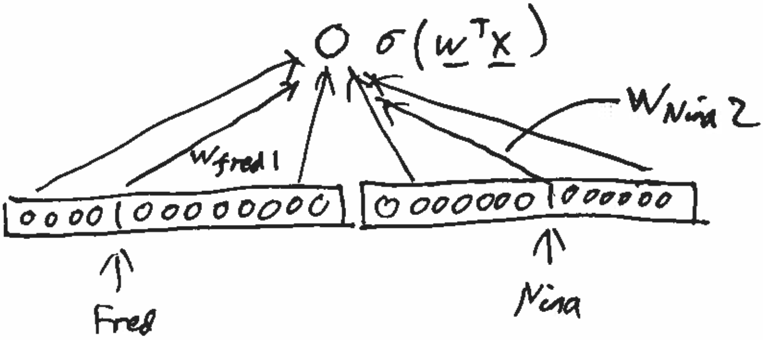
Training data

How might we predict winner given players?

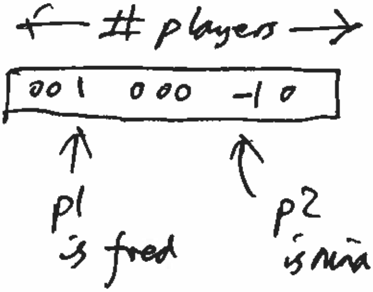
Can we rank players, best → worst?

Think of simple approaches first.

Each player "skill parameter"



$$p(\underbrace{y=1}_{\text{player 1 wins (fred)}} \mid \underline{x}, \underline{w}) = \sigma(\underbrace{w_{fred\,1}}_{\text{"skill fred"}} + \underbrace{w_{nina\,2}}_{\text{"Badness Nina"}})$$



$$\sigma(w_{fred} - w_{nina} + b)$$

Approx. Normalizer $p(D|M)$

$$p(\underline{w}|D) = \frac{p(\underline{w}, D)}{P(D)} \approx \mathcal{N}(\underline{w}; \underline{w}^*, H^{-1})$$

$$= \frac{|H|^{1/2}}{(2\pi)^{D/2}} e^{-1/2(\underline{w}-\underline{w}^*)^T H (\underline{w}\ldots}$$

Evaluate approx. at $\underline{w} = \underline{w}^*$:

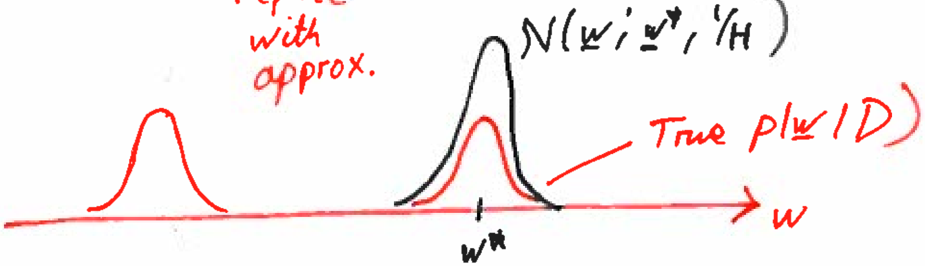
$$\frac{p(\underline{w}^*, D)}{p(D)} \approx \frac{|H|^{1/2}}{(2\pi)^{D/2}}$$

parameters

Training data

$$P(D) = \frac{p(\underline{w}^*, D)}{P(\underline{w}^*|D)} \approx \frac{p(\underline{w}^*, D)}{|H|^{1/2}} (2\pi)^{D}$$

replace with approx.

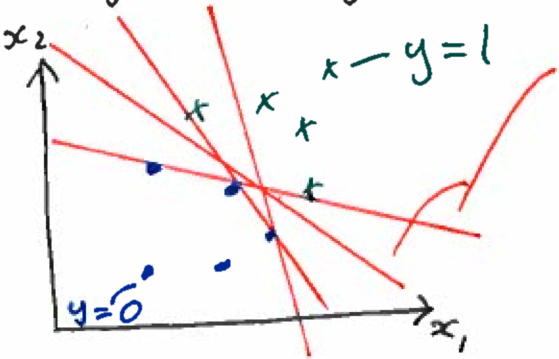


$P(D)$ from Laplace approximation:

A) Too Big; B) Too Small; C) Correct; Z) ???

$$p(D) = \int p(D, \underline{w})\, d\underline{w}$$
$$= \int p(D \mid \underline{w})\, p(\underline{w})\, d\underline{w}$$

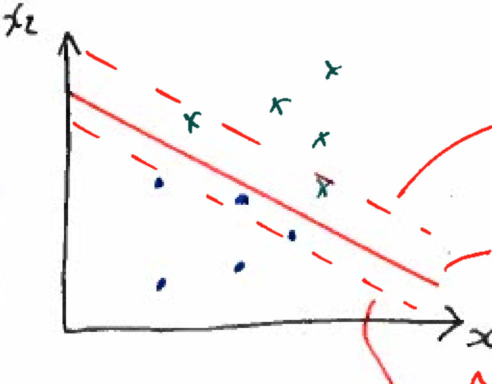
Bayesian Logistic Regression



Decision boundaries:

$p(y=1 \mid \underline{x}, \underline{w}) = \sigma(\underline{w}^T\underline{x}) = \frac{1}{2}$

for different plausible $\underline{w}$

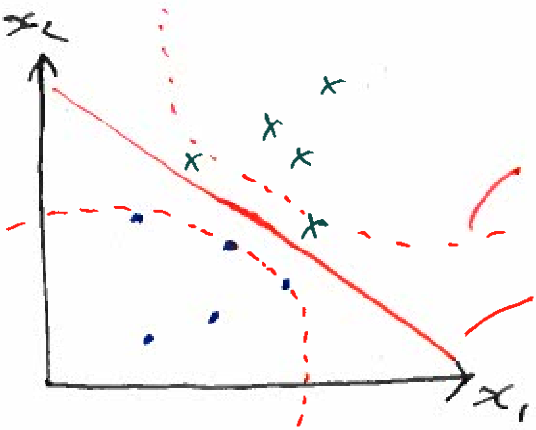


$\sigma(\hat{\underline{w}}^T\underline{x}) = 0.9$

$\sigma(\hat{\underline{w}}^T\underline{x}) = 0.5$

$\sigma(\hat{\underline{w}}^T\underline{x}) = 0.1$

Predictive distribution for a single fitted $\hat{\underline{w}}$, eg L2 regularised fit



$p(y=1 \mid \underline{x}, D) = 0.9$

$p(y=1 \mid \underline{x}, D) = 1/2$

Training data

Plausible weights described by Posterior

$$p(\underline{w} \mid D, M) = \frac{p(D \mid \underline{w}, M)\, p(\underline{w} \mid M)}{p(D \mid M)}$$

Train $\{\underline{x}^{(n)}, y^{(n)}\}$

$p(D \mid M) = \int p(D \mid \underline{w})\, p(\underline{w})\, d$

M, model choices, hyperparameters, basis $f^{\underline{n}}$'s
(Often miss out)

Likelihood

"known" here

$$P(D \mid \underline{w}) = \prod_n \cancel{p(\underline{x}^{(n)} \mid \underline{w})}\, \underbrace{p(y^{(n)} \mid \underline{w}, \underline{x}^{(n)})}_{\sigma(\underline{w}^T \underline{x}^{(n)}(2y^{(n)} - 1))}$$

Prior: $p(\underline{w}) = \mathcal{N}(\underline{w}; \underline{0}, \sigma_w^2 \mathbb{I})$

Marginal Likelihood $p(D)$ or $p(D \mid M)$ can compare models

Predictions

$$p(y=1 \mid \underline{x}, D) = \int p(y=1, \underline{w} \mid \underline{x}, D)\, d\underline{w}$$

test output, test input

$$= \int \underbrace{p(y=1 \mid \underline{w}, \underline{x}, \cancel{D})}_{\sigma(\underline{w}^T \underline{x})}\, \underbrace{p(\underline{w} \mid \cancel{\underline{x}}, D)}_{\text{posterior}}\, d\underline{w}$$

Predictions for logistic regression

$$p(y=1 \mid \underline{x}, D) = \int \sigma(\underline{w}^T\underline{x})\, \underbrace{N(\underline{w}; \underline{w}^*, H^{-1})}_{\text{Laplace approx.}}\, d\underline{w}$$

$$= \mathbb{E}_{N(\underline{w};\underline{w}^*, H^{-1})}\left[\sigma(\underbrace{\underline{w}^T\underline{x}}_{\text{scalar}})\right]$$

(Could do Monte Carlo)

Average under "activation" $\underline{w}^T\underline{x} = a$

$$= \mathbb{E}_{\underbrace{p(a)}_{N(a;\, \underline{w}^{*T}\underline{x},\, \underbrace{\underline{x}^T H^{-1} \underline{x}}_{\text{Homework: check!}})}} [\sigma(a)]$$

$$= \int \sigma(a)\, N(a; \underline{w}^{*T}\underline{x}, \underline{x}^T H^{-1} \underline{x})\, da$$

Could solve numerically

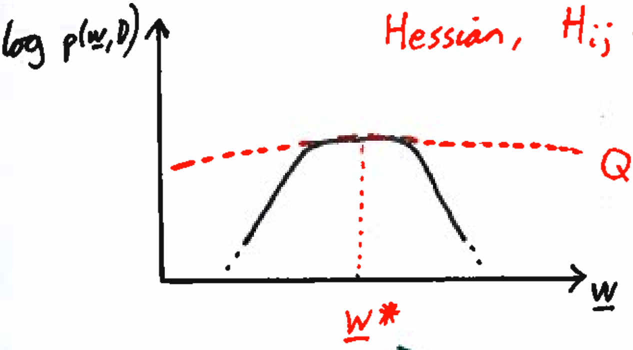
$$\approx \sigma(\kappa\, \underline{w}^{*T}\underline{x})$$

$$\kappa = \frac{1}{\sqrt{1 + \pi/8\, \underline{x}^T H^{-1} \underline{x}}}$$

Bishop p220
Murphy §8.4.4.2

Laplace Approximation

Hessian, $H_{ij} = \frac{\partial -\log p(\underline{w}, D)}{\partial w_i \, \partial w_j}$



Quadratic fit,

tiny curvature

Found with optimizer

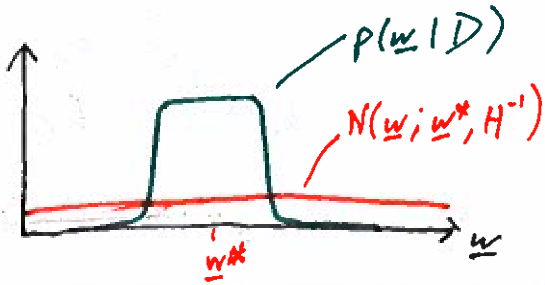
Posterior

$$p(\underline{w} | D) \approx N(\underline{w}; \underline{w}^*, H^{-1})$$

Marginal Likelihood

$$p(D) = \frac{p(\underline{w}^*, D)}{p(\underline{w}^* | D)} \approx \frac{p(\underline{w}^*, D)}{N(\underline{w}^*; \underline{w}^*, H^{-1})}$$

Areas under both curves = 1



Quiz

Is p(D) approx.

A) Too big

B) Too small

C) ≈ Correct

Z) ???