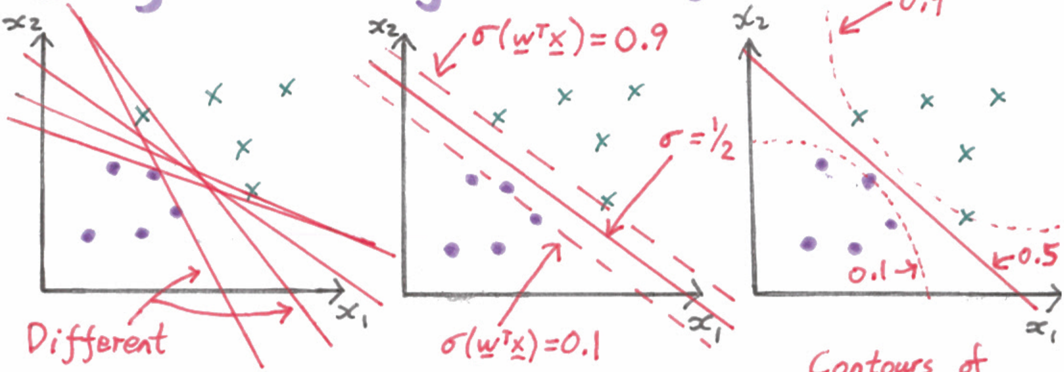
Bayesian Logistic Regression



Different plausible decision boundaries

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

for different $\underline{w} \sim$ posterior

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

Contours for single fitted $\underline{w}$

Contours of $p(y=1|\underline{x}, \mathcal{D})$

Training data

Posterior

Likelihood: Large product sigmoids

$$p(\underline{w}|\mathcal{D}, M) = \frac{\overbrace{P(\mathcal{D}|\underline{w}, M)}^{\text{Likelihood}}\, p(\underline{w}|M)}{P(\mathcal{D}|M)}$$

Model, Hyperparameters, Basis functions

Marginal Likelihood

$\int P(\mathcal{D}|\underline{w})\, p(\underline{w})\, d\underline{w}$

Predictions

Training data

$$P(y|\underline{x}, \mathcal{D}) = \int p(y|\underline{x},\underline{w})\, \underbrace{p(\underline{w}|\mathcal{D})}_{\text{Posterior: not Gaussian}}\, d\underline{w}$$

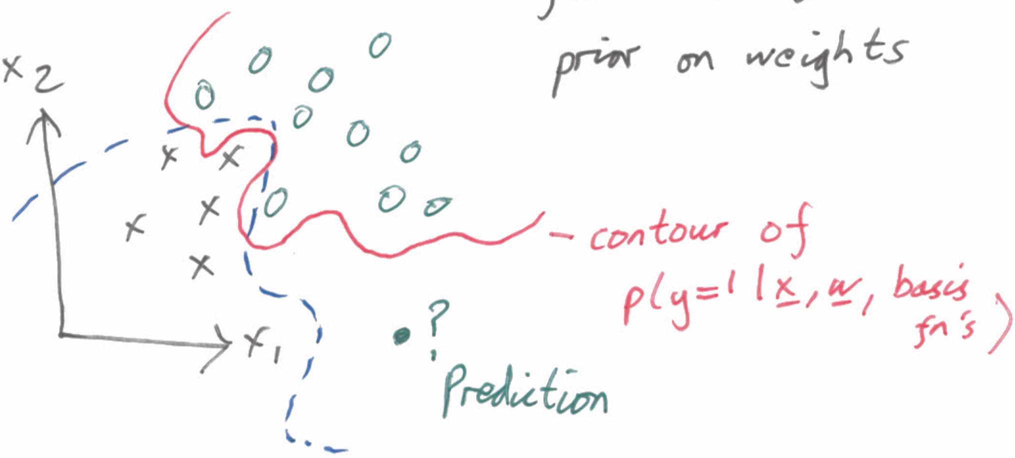
Test input

Posterior $p(\underline{w} \mid D)$

within some model

fixed basis functions

prior on weights



Marginal Likelihood $p(D)$

Importance Sampling (special case, using prior)

$$P(y|\underline{x},D) = \int p(y|\underline{x},\underline{w})\, \frac{P(D|\underline{w})}{P(D)}\, p(\underline{w})\, d\underline{w}$$

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

$$\approx \cancel{\frac{1}{S}} \sum_{s=1}^{S} p(y|\underline{x},\underline{w}^{(s)})\, \frac{P(D|\underline{w}^{(s)})}{\cancel{\frac{1}{S}} \sum_{s'} p(D|\underline{w}^{(s')})}$$

$\approx$

$\underline{w}^{(s)} \sim$ prior $p(\underline{w})$

Laplace Approximation

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

$\underline{w}^* = \underset{\underline{w}}{\operatorname{argmin}}\, E(\underline{w})$ $\qquad$ $H_{ij} = \dfrac{\partial^2 E}{\partial w_i \partial w_j}$

"Energy", $E = -\log p(\underline{w}, D) + \text{const wrt } \underline{w}$

If posterior is Gaussian, is correct.

Otherwise an approximation ("wrong"!)

Predictions for logistic regression

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

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

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

(Could do Monte Carlo)

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

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

$$= \int \sigma(a)\, N(a; \underline{w}^{*\top}\underline{x}, \underbrace{\underline{x}^\top H^{-1} \underline{x}}_{\text{scalar}})\, da$$

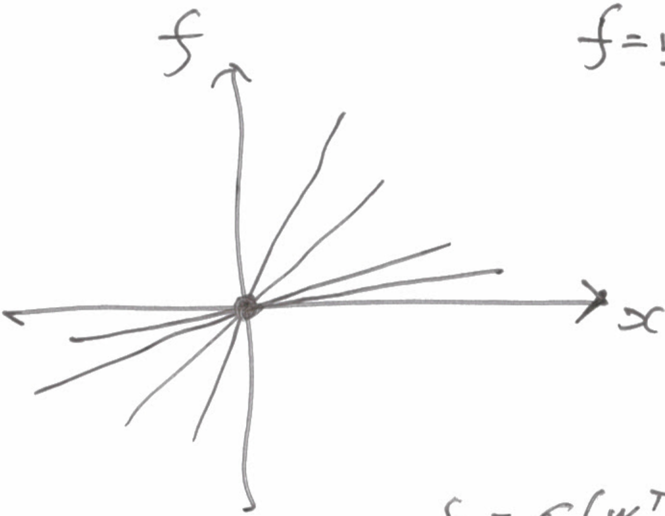
Could solve numerically

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

Murphy §8.4.4.2

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

$$f = \underline{w}^T \underline{x}$$



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

Approx. Normalizer $p(D)$

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

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

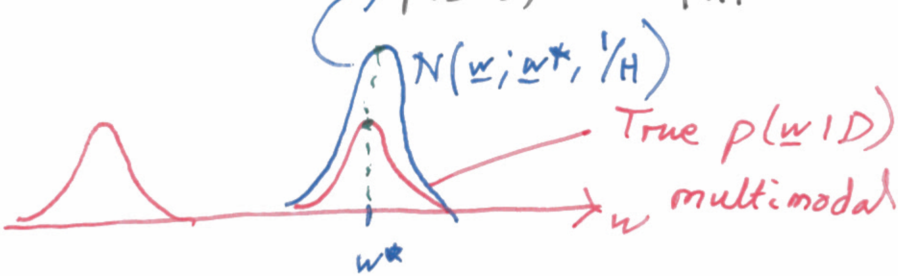
Evaluate approx. at $\underline{w} = \underline{w}^*$:

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

Training set

\# parameter

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



$P(D)$ approx with Laplace

A) Too Big, B) Too Small C) Correct,

Z) ????

Sketch logistic Regression Posterior for 1 data point

$$p(w) = \mathcal{N}(w; 0, 1)$$

$x = -20$
$y = +1$

$$p(w \mid D) \propto \mathcal{N}(w; 0, 1)\, \sigma(10 - 20w)$$

know bias = 10