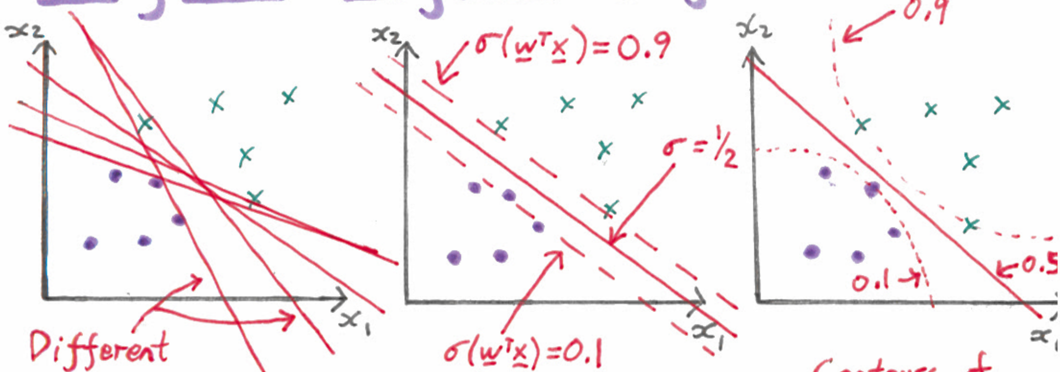
Bayesian Logistic Regression



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

Contours for single fitted $\underline{w}$

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

$\uparrow$ Training data

Posterior

Likelihood: Large product sigmoids

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

$\uparrow$ Model, Hyperparameters, Basis functions

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

Predictions

Training data

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

$\uparrow$ Test input

Monte Carlo

$$p(y \mid \underline{x}, D) = \mathbb{E}_{p(\underline{w} \mid D)}\left[p(y \mid \underline{x}, \underline{w}) \right]$$

$$\approx \frac{1}{S} \sum_{s=1}^{S} p(y \mid \underline{x}, \underline{w}^{(s)}), \quad \underbrace{\underline{w}^{(s)} \sim p(\underline{w} \mid D)}$$

How? Approximately with "MCMC" (not this course)

Importance Sampling

$$\int g(x)\, p(x)\, dx = \int \left[g(x) \frac{p(x)}{q(x)} \right] q(x)\, dx$$

$$\mathbb{E}_p[g(x)] = \mathbb{E}_q[\quad \ldots \quad] \quad \left[\begin{array}{l} q(x) \neq 0 \\ \text{If } p(x) \neq 0 \end{array} \right]$$

$$= \frac{1}{S} \sum_{s=1}^{S} [\ldots], \quad x \sim q$$

For logistic regression predictions

$\underline{w} \sim p(\underline{w})$ prior (maybe Gaussian)

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

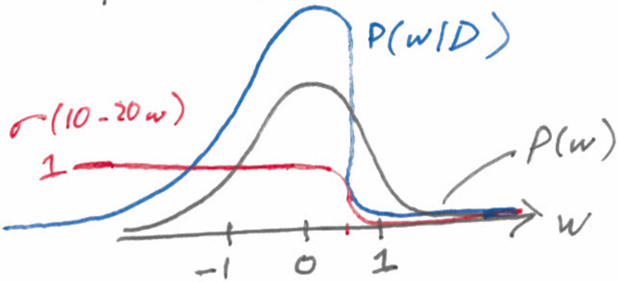
$$\underline{w}^{(s)} \sim p(\underline{w})$$

Sketch Posterior for one data point

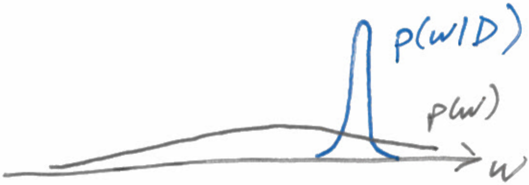
$$p(w) = N(w; 0, 1) \qquad x = -20,\ y = +1$$

$$p(w|D) \propto N(w; 0, 1)\, \sigma(10 - 20w)$$

↑ know bias = 10



In the notes: $N = 500$ posterior



Laplace Approximation

Fits a Gaussian to $p(\underline{w}|D)$

- Matching mode
- Match curvature (2nd derivative) of $\log p(\underline{w}|D)$

$$\underline{w}^* = \underset{\underline{w}}{\operatorname{argmax}}\ p(\underline{w}|D)$$ (MAP parameters)

Gaussian prior → L2 regularization

$$= \underset{\underline{w}}{\operatorname{argmin}}\ \underbrace{-\log p(\underline{w}|D)}$$ Numerical optimization

"Energy" Instead.

$$E(\underline{w}) = -\log p(\underline{w}, D)$$ neg. log posterior up to constant

Find minimum w^*, then find curvature:

1D $$H = \left.\frac{\partial^2 E}{\partial w^2}\right|_{w=w^*}$$

In general Hessian $$H_{ij} = \left.\frac{\partial^2 E}{\partial w_i \partial w_j}\right|_{\underline{w}=\underline{w}^*}$$

Compare to Gaussian

"Energy" = − log prob + const. wrt w

$$E_{\mathcal{N}(w;\mu,\sigma^2)}(w) = \frac{(w-\mu)^2}{2\sigma^2}$$

Minimum: $w^* = \mu$

$$H = \frac{1}{\sigma^2} \Rightarrow \sigma^2 = \frac{1}{H}$$

$$E_{N(\underline{w};\,\underline{\mu},\,\Sigma)} = \tfrac{1}{2}(\underline{w}-\underline{\mu})^T \Sigma^{-1} (\underline{w}-\underline{\mu})$$

$$\underline{w}^* = \underline{\mu}$$

$$H = \Sigma^{-1} \Rightarrow \Sigma = H^{-1}$$

Laplace Approximation:

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

Approx. Normalizer $p(D)$

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

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

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

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

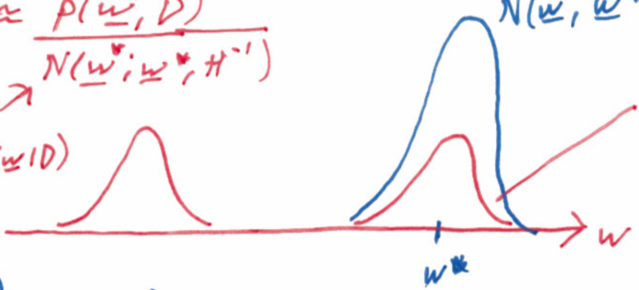
parameters

Training data

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

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

Should be $p(\underline{w} \mid D)$



$P(D)$ approx:

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

Z) ????