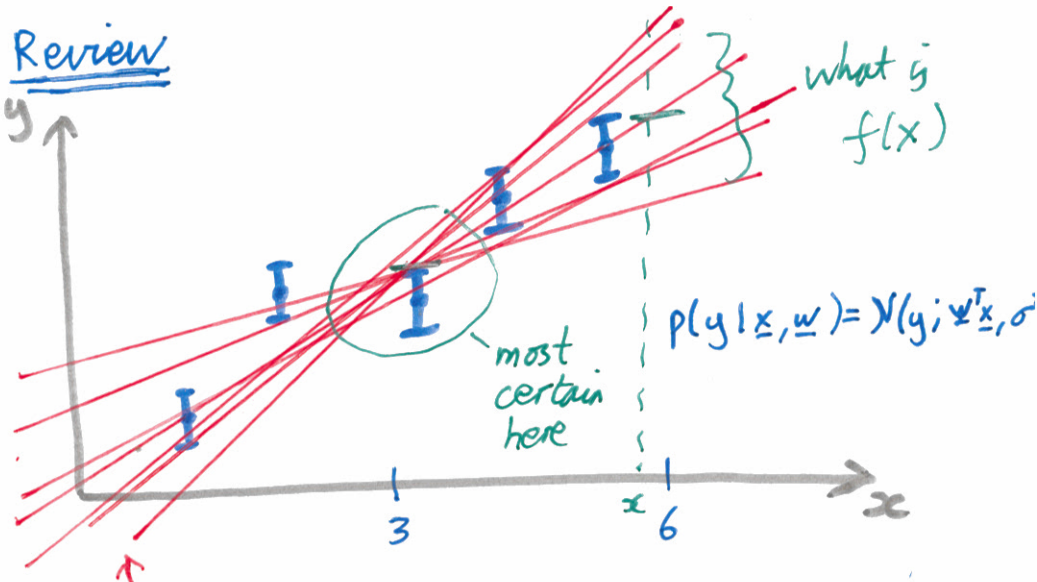
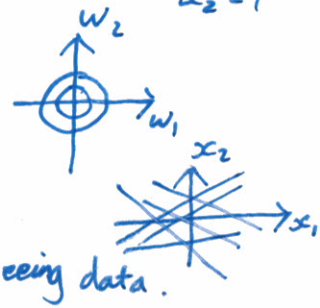
Review



Bayesian Inference

$x_2 = 1$



Prior $p(\underline{w}) = N(\underline{w}; 0, \sigma_{prior}^2 I)$

Many functions plausible before seeing data.

Posterior

$$p(\underline{w}|D) = N(\underline{w}; \underline{w}_N, V_N)$$
$$\propto p(\underline{w})\, p(\underline{y}|X, \underline{w})$$

Prediction

$$p(f(\underline{x})|D) = N(f; \underline{w}_N^T \underline{x}, \underline{x}^T V_N \underline{x})$$

$$p(y|\underline{x}, D) = N(y; \underline{w}_N^T \underline{x}, \underline{x}^T V_N \underline{x} + \sigma^2)$$

Question

Uncertainty $\underline{x}^T V_N \underline{x}$ grows with $\underline{x}$.

Why in the figure is most certain region at $x > 0$ (around $x = 3$)?

How is that possible in the maths?

Ideas

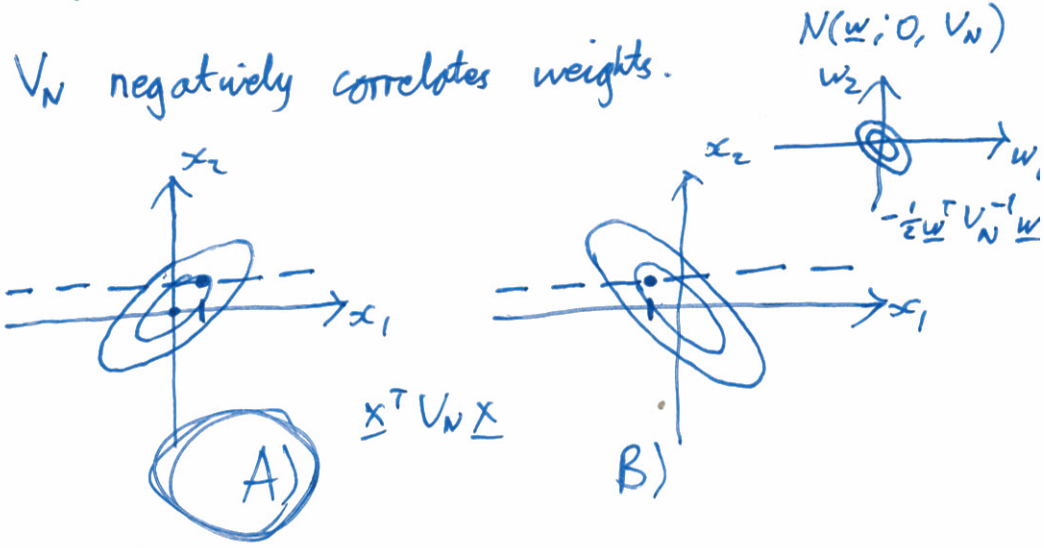
V_N could be a function $\underline{x}$?

Not a function of where we test.

Quadratic Equation

$(x-t)^2$ min is at t not zero.

V_N negatively correlates weights.



C) other

Z) ???

Overfitting

Bayesian don't fit so can't "over fit"

We don't fit

$$\hat{\underline{w}} = \underset{\underline{w}}{\operatorname{argmin}}\ \text{cost}(\underline{w})$$

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

Predictions sum / integrate.

"Underfitting"

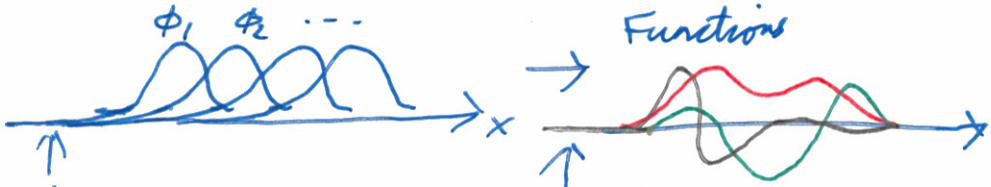
Over-simple models → Over confidence

Residuals tell us model is wrong.

Model checking.

With Basis functions

Function is a combination of



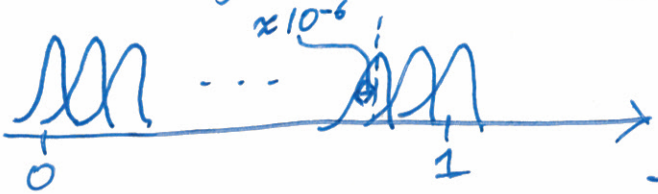
$\underline{\phi} \approx \underline{0}$

We know $f \approx 0$ here

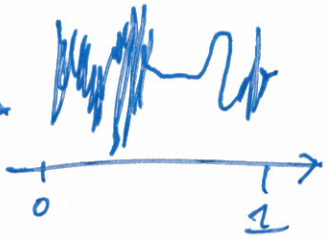
$\underline{x}^T V_N \underline{x} \Rightarrow \underline{\phi}^T V_N \underline{\phi}$

Big Models cause problems: extreme example

10^6 basis functions



Can model



Or basis f^'s like a top hat

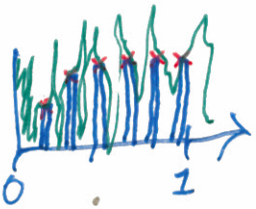
10^6 of those

Represent any piecewise const f^ with 10^6 pieces

$p(w_i) = \mathcal{N}(w_i; 0, \sigma_f^2)$

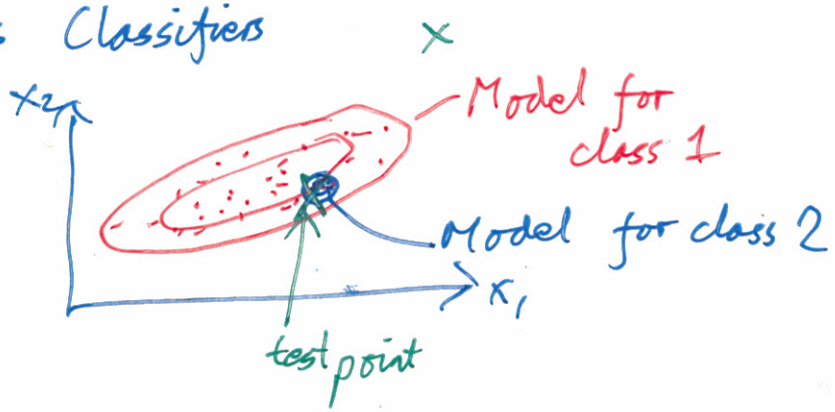
What's posterior?

$p(\underline{w} | \underline{y}, X) \propto \underbrace{p(\underline{y} | \underline{w}, X)}_{\prod_n p(y_n | \underline{w}, X)} p(\underline{w})$



Probabilistic Model Choice

Bayes Classifiers



Regression

Regression model

$$p(\underline{y} \mid X, M) = \int p(\underline{y} \mid X, \underline{w}, M) P(\underline{w} \mid M) \, d\underline{w}$$

Likelihood of Model.