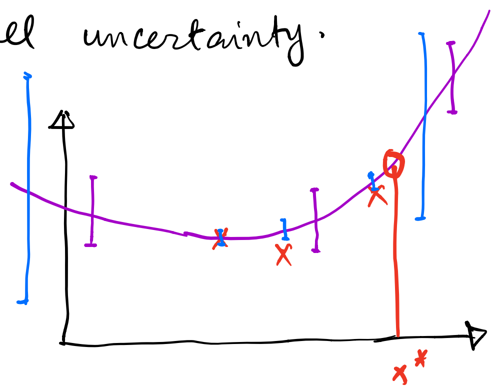
Bayesian Reasoning

Good news: Rare disease. 1/10000 have it.

Bad news: You tested positive for a terminal disease. Test is 99%..

Model uncertainty.



Bayes Rule: $$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$$

$$P(A \cap B) = \underbrace{P(A|B)}_{\text{Conditional}}\ \underbrace{P(B)}_{\text{marginal}} = P(B|A)\,P(A) = \underbrace{P(A,B)}_{\text{Joint Prob.}}$$

① $\int P(A,B)\, dA\, dB = 1$

② $\int P(A|B)\, dA = 1$

just the prob. of A. B is fixed.

③ $\int P(A)\, dA = 1$

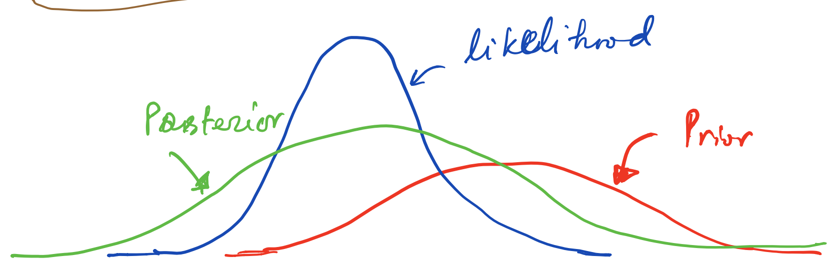
④ $\int P(A,B)\, dB = P(A)$ Integrating out B (a.k.a. marginalization)

Learning and Bayesian Reasoning

h : hypothesis

d : data

$$\boxed{p(h|d)} = \frac{\boxed{p(d|h)}\,\boxed{p(h)}}{\boxed{\sum_{h'\in H} p(d|h')\,dh'}} \quad \left(= p(d)\right)$$



Very difficult to compute.

① Test is 99% accurate:

$$p(T=1 \mid D=1) = 0.99$$

$$p(T=0 \mid D=0) = 0.99$$

② $p(D=1) = 0.0001$

$p(D=0) = 0.9999$

$$p(D=1 \mid T=1) = \frac{p(T=1|D=1)\,p(D=1)}{p(T=1|D=0)\,p(D=0) + p(T=1|D=1)\,p(D=1)}$$

$$= \frac{(0.99)(0.0001)}{(1-0.99)(0.9999) + (0.99)(0.0001)}$$

$= 0.0098$

Bayesian Linear Regression

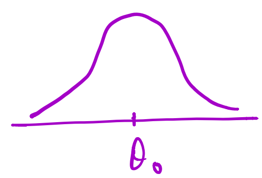
Parameters: θ

Data: D

$$P(\theta \mid D) = \frac{P(D \mid \theta)\, P(\theta)}{P(D)} \propto P(D \mid \theta)\, P(\theta)$$

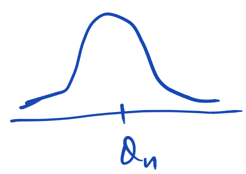
$P(D)$ ← Normalizing factor

1. Likelihood: $\mathcal{N}(y \mid X\theta, \sigma^2 I_n)$



2. Prior: $\mathcal{N}(\theta \mid \theta_0, V_0)$

θ_0: mean, V_0: variance



3. Posterior: $\mathcal{N}(\theta \mid \theta_n, V_n)$

θ_n: ? V_n: ?

$$P(\theta \mid x, y, \sigma^2) \propto \boxed{\mathcal{N}(y \mid X\theta, \sigma^2 I_n)}\,\boxed{\mathcal{N}(\theta \mid \theta_0, V_0)}$$

Removed equality '=' with proportionality '∝' by getting rid of the denominator term.

$$P(\theta \mid x, y, \sigma^2) \propto e^{-\frac{1}{2\sigma^2}(Y - X\theta)^T (Y - X\theta)} \; e^{-\frac{1}{2}(\theta - \theta_0)^T V_0^{-1} (\theta - \theta_0)}$$

$\in \mathbb{R}^3$

Aside: $y = \theta_0 + \theta_1 x + \theta_2 x^2$

$\Longrightarrow \quad \theta = [\theta_0 \quad \theta_1 \quad \theta_2]$

$$\propto e^{-\frac{1}{2}\{(Y - X\theta)^T \sigma^{-2} (Y - X\theta) + (\theta - \theta_0)^T V_0^{-1} (\theta - \theta_0)\}}$$

$$\propto e^{-\frac{1}{2}\{\sigma^{-2} Y^T Y - 2\sigma^{-2} Y^T X\theta + \boxed{\sigma^{-2}\theta^T X^T X\theta + \theta^T V_0^{-1}\theta} - 2\theta_0^T V_0^{-1}\theta + \theta_0^T V_0^{-1}\theta_0\}}$$

x_1, y_1

$\vdots$

x_n, y_n

$$X \begin{bmatrix} 1 & x_1 & x_1^2 \\ \vdots & & \\ 1 & x_n & x_n^2 \end{bmatrix} \qquad Y \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

$$\sigma^{-2}\theta^T X^T X\theta + \theta^T V_0^{-1}\theta$$

$$= \theta^T (X^T (\sigma^2 I)^{-1} X + V_0^{-1})\theta$$

$$= \theta^T V_n^{-1}\theta$$

$$= e^{-\frac{1}{2}\{\text{const} + \theta^T V_n^{-1}\theta - 2\left(\frac{Y^T X}{\sigma^2} + \theta_0^T V_0^{-1}\right)\theta\}}$$

$$= e^{-\frac{1}{2}\{\text{const} + \theta^T V_n^{-1}\theta - 2\theta_n^T V_n^{-1}\theta + 2\theta_n^T V^{-1}\theta - 2\left(\frac{Y^T X}{\sigma^2} + \theta_0^T V_0^{-1}\right)\theta\}}$$

$$= e^{-\frac{1}{2}\{\text{const2} + \boxed{(\theta - \theta_n)^T V_n^{-1} (\theta - \theta_n)} + 2\left(\theta_n^T V_n^{-1} - \frac{Y^T X}{\sigma^2} - \theta_0^T V_0^{-1}\right)\theta\}}$$

don't care upto prop.

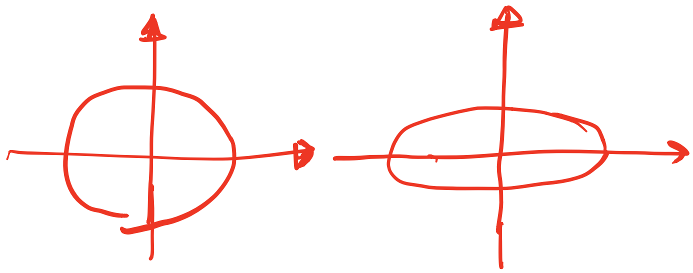
Ideally $= 0$ to make a Gaussian.

$$\theta_n^T V_n^{-1} - \frac{y^T x}{\sigma^2} - \theta_0^T V_0^{-1} = 0$$

$$\Rightarrow \theta_n^T = \left(\frac{y^T x}{\sigma^2} + \theta_0^T V_0^{-1} \right) V_n$$

Assume: θ_n , V_n^{-1}

What if $\theta_0 = 0$, $V_0 = \tau_0^2 I_d$



$$\mathcal{N}(\theta \mid \theta_n V_n) \propto e^{-\frac{1}{2}\{ (\theta - \theta_n)^T V_n^{-1} (\theta - \theta_n) \}}$$

$$= \left| 2\pi |V_n| \right|^{-1/2}$$

Theorem of Gaussians KM Book. Ch. 4

Bayesian Regression: Bayesian vs. ML Prediction

$$\theta_n = (\lambda I_d + X^T X)^{-1} X^T y$$

$$V_n = \sigma^2 (\lambda I_d + X^T X)^{-1}$$

$$\lambda = \sigma^2 / \tau_0^2$$

Given some new data point x^*

$$p(y \mid x^*, D, \sigma^2) = \int \mathcal{N}(y \mid x^{*T}\theta, \sigma^2)\, \mathcal{N}(\theta \mid \theta_n, V_n)\, d\theta$$

$$= \mathcal{N}(y \mid x^{*T}\theta_n, \sigma^2 \underbrace{+ x^{*T} V_n x^*}_{?})$$

ML:

$$p(y \mid x^*, D, \sigma^2) = \mathcal{N}(y \mid x^{*T}\theta_{ML}, \sigma^2)$$