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: \[ P(A|B) = \frac{P(B|A)P(A)}{P(B)} \]

\[ P(A \cap B) = P(A|B)P(B) = P(B|A)P(A) = P(A,B) \]

\[ \downarrow \text{Conditional marginal} \]

\[ \text{Joint Prob.} \]

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

\[ p(h \mid d) = \frac{p(d \mid h) p(h)}{\sum_{h'} p(d \mid h') p(h')} \]

(= p(d))

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. \( p(D=1) = 0.0001 \)
   \[ p(D=0) = 0.9999 \]
   \[ p(D=1 \mid T=1) = \frac{p(T=1 \mid D=1) p(D=1)}{p(T=1 \mid D=0) p(D=0) + p(T=1 \mid D=1) p(D=1)} \]
   \[ = \frac{(0.99)(0.0001)}{(1-0.99)(0.9999) + (0.99)(0.0001)} \]
Bayesian Linear Regression

Parameters: $\theta$

Data: $D$

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

Normalization factor

1. Likelihood: $\mathcal{N}(y | x^T \theta, \sigma^2 I_n)$

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

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

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

Removed equality '=' with proportionality '$\propto$' by getting rid of the denominator term.

$$P(\theta | x, y, \sigma^2) \propto e^{-\frac{1}{2\sigma^2}(y-x^T\theta)^T(y-x^T\theta) - \frac{1}{2} (\theta - \theta_0)^T V^{-1}_0 (\theta - \theta_0)} \in \mathbb{R}^3$$
Aside: \[ y = \theta_0 + \theta_1 x + \theta_2 x^2 \]
\[ \Rightarrow \theta = [\theta_0, \theta_1, \theta_2] \]

\[ -\frac{1}{2} \left\{ (y-x^T \theta)^T \Sigma^{-1} (y-x^T \theta) + (\theta - \theta_0)^T V_\theta^{-1} (\theta - \theta_0) \right\} \]

\[ = -\frac{1}{2} \left\{ \theta^T \Sigma^{-1} \theta - 2 \theta^T \Sigma^{-1} x + \theta^T \Sigma^{-1} \theta_0 + \theta_0^T \Sigma^{-1} \theta_0 - 2 \theta_0^T V_\theta^{-1} \theta \right\} + \theta_0^T V_\theta^{-1} \theta_0 \]

\[ \alpha \in \mathbb{E} \left\{ X \begin{bmatrix} 1 & x_1 & x_1^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \right\} \]

\[ = \begin{bmatrix} \theta \end{bmatrix}^T \begin{bmatrix} \alpha \end{bmatrix} \]

\[ = \begin{bmatrix} \alpha \end{bmatrix}^T \begin{bmatrix} \theta \end{bmatrix} - 2 \begin{bmatrix} \alpha \end{bmatrix}^T \begin{bmatrix} \theta_0 \end{bmatrix} + \begin{bmatrix} \theta_0 \end{bmatrix}^T \begin{bmatrix} \theta \end{bmatrix} \]

\[ = \begin{bmatrix} \theta \end{bmatrix}^T \begin{bmatrix} \alpha \end{bmatrix} - 2 \begin{bmatrix} \alpha \end{bmatrix}^T \begin{bmatrix} \theta_0 \end{bmatrix} + \begin{bmatrix} \theta_0 \end{bmatrix}^T \begin{bmatrix} \theta \end{bmatrix} \]

\[ = \begin{bmatrix} \theta \end{bmatrix}^T \begin{bmatrix} \alpha \end{bmatrix} - 2 \begin{bmatrix} \alpha \end{bmatrix}^T \begin{bmatrix} \theta_0 \end{bmatrix} + \begin{bmatrix} \theta_0 \end{bmatrix}^T \begin{bmatrix} \theta \end{bmatrix} \]

\[ = \begin{bmatrix} \theta \end{bmatrix}^T \begin{bmatrix} \alpha \end{bmatrix} - 2 \begin{bmatrix} \alpha \end{bmatrix}^T \begin{bmatrix} \theta_0 \end{bmatrix} + \begin{bmatrix} \theta_0 \end{bmatrix}^T \begin{bmatrix} \theta \end{bmatrix} \]

\[ = \begin{bmatrix} \theta \end{bmatrix}^T \begin{bmatrix} \alpha \end{bmatrix} - 2 \begin{bmatrix} \alpha \end{bmatrix}^T \begin{bmatrix} \theta_0 \end{bmatrix} + \begin{bmatrix} \theta_0 \end{bmatrix}^T \begin{bmatrix} \theta \end{bmatrix} \]

\[ = \begin{bmatrix} \theta \end{bmatrix}^T \begin{bmatrix} \alpha \end{bmatrix} - 2 \begin{bmatrix} \alpha \end{bmatrix}^T \begin{bmatrix} \theta_0 \end{bmatrix} + \begin{bmatrix} \theta_0 \end{bmatrix}^T \begin{bmatrix} \theta \end{bmatrix} \]

\[ = \begin{bmatrix} \theta \end{bmatrix}^T \begin{bmatrix} \alpha \end{bmatrix} - 2 \begin{bmatrix} \alpha \end{bmatrix}^T \begin{bmatrix} \theta_0 \end{bmatrix} + \begin{bmatrix} \theta_0 \end{bmatrix}^T \begin{bmatrix} \theta \end{bmatrix} \]

\[ = \begin{bmatrix} \theta \end{bmatrix}^T \begin{bmatrix} \alpha \end{bmatrix} - 2 \begin{bmatrix} \alpha \end{bmatrix}^T \begin{bmatrix} \theta_0 \end{bmatrix} + \begin{bmatrix} \theta_0 \end{bmatrix}^T \begin{bmatrix} \theta \end{bmatrix} \]

\[ = \begin{bmatrix} \theta \end{bmatrix}^T \begin{bmatrix} \alpha \end{bmatrix} - 2 \begin{bmatrix} \alpha \end{bmatrix}^T \begin{bmatrix} \theta_0 \end{bmatrix} + \begin{bmatrix} \theta_0 \end{bmatrix}^T \begin{bmatrix} \theta \end{bmatrix} \]

\[ = \begin{bmatrix} \theta \end{bmatrix}^T \begin{bmatrix} \alpha \end{bmatrix} - 2 \begin{bmatrix} \alpha \end{bmatrix}^T \begin{bmatrix} \theta_0 \end{bmatrix} + \begin{bmatrix} \theta_0 \end{bmatrix}^T \begin{bmatrix} \theta \end{bmatrix} \]

\[ = \begin{bmatrix} \theta \end{bmatrix}^T \begin{bmatrix} \alpha \end{bmatrix} - 2 \begin{bmatrix} \alpha \end{bmatrix}^T \begin{bmatrix} \theta_0 \end{bmatrix} + \begin{bmatrix} \theta_0 \end{bmatrix}^T \begin{bmatrix} \theta \end{bmatrix} \]

\[ = \begin{bmatrix} \theta \end{bmatrix}^T \begin{bmatrix} \alpha \end{bmatrix} - 2 \begin{bmatrix} \alpha \end{bmatrix}^T \begin{bmatrix} \theta_0 \end{bmatrix} + \begin{bmatrix} \theta_0 \end{bmatrix}^T \begin{bmatrix} \theta \end{bmatrix} \]
\[ \Theta_n^T V_n^{-1} - \frac{y^T x}{\sigma^2} - \Theta_o^T V_o^{-1} = 0 \]

\[ \Rightarrow \Theta_n^T = \left( \frac{y^T x}{\sigma^2} + \Theta_o^T V_o^{-1} \right) V_n \]

Assume: \( \Theta_n, V_n^{-1} \)

What if \( \Theta_o = 0, V_o = \sigma_o^2 I_d \)

\[ N(\Theta | \Theta_n, V_n) \propto e^{-\frac{1}{2} \frac{1}{2} (\Theta - \Theta_n)^T V_n^{-1} (\Theta - \Theta_n)} \]

\[ = \frac{1}{2 \pi |V_n|} \left| \frac{1}{2} \right|^{\frac{1}{2}} \]

Theorem of Gaussians KM Book Ch. 4
Bayesian Regression: Bayesian vs. ML Predictor

\[ \theta_n = (\lambda I_d + X^T X)^{-1} X^T \gamma \]

\[ \nu_n = \sigma^2 (\lambda I_d + X^T X)^{-1} \]

\[ \lambda = \sigma^2 / \gamma_o^2 \]

Given some new data point \( x^* \)

\[ p(y | x^*, D, \sigma^2) = \int \mathcal{N}(y | x^T \theta, \sigma^2) \mathcal{N}(\theta | \theta_n, \nu_n) \, d\theta \]

\[ = \mathcal{N}(y | x^T \theta_n, \sigma^2 + x^T \nu_n x^*) \]

ML:

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