The Rules of Probability

- Sum Rule
  
  \[ p(X) = \sum_Y p(X, Y) \]

- Product Rule
  
  \[ p(X, Y) = p(Y | X)p(X) \]
\[ p(X|Y = 1) \]

\[ p(X) \]

\[ p(X, Y) \]

\[ p(Y) \]
Bayes’ Theorem

\[ p(Y|X) = \frac{p(X|Y)p(Y)}{p(X)} \]

\[ p(X) = \sum_{Y} p(X|Y)p(Y) \]

posterior \propto \text{likelihood} \times \text{prior}
Probability Theory

Ex: Apples and Oranges

40% 60%
The Gaussian Distribution

\[ \mathcal{N}(x|\mu, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp \left\{ -\frac{1}{2\sigma^2} (x - \mu)^2 \right\} \]

\[ \mathcal{N}(x|\mu, \sigma^2) > 0 \]

\[ \int_{-\infty}^{\infty} \mathcal{N}(x|\mu, \sigma^2) \, dx = 1 \]
Gaussian Mean and Variance

\[ E[x] = \int_{-\infty}^{\infty} \mathcal{N}(x|\mu, \sigma^2) x \, dx = \mu \]

\[ E[x^2] = \int_{-\infty}^{\infty} \mathcal{N}(x|\mu, \sigma^2) x^2 \, dx = \mu^2 + \sigma^2 \]

\[ \text{var}[x] = E[x^2] - E[x]^2 = \sigma^2 \]
Gaussian Parameter Estimation

\[ p(x|\mu, \sigma^2) = \prod_{n=1}^{N} \mathcal{N}(x_n|\mu, \sigma^2) \]

Likelihood function

notes
Properties of $\sigma^2_{ML}$ and $\mu_{ML}$

$$\mathbb{E}[\mu_{ML}] = \mu$$

$$\mathbb{E}[\sigma^2_{ML}] = \left( \frac{N - 1}{N} \right) \sigma^2$$

$$\tilde{\sigma}^2 = \frac{N}{N - 1} \sigma^2_{ML}$$

$$= \frac{1}{N - 1} \sum_{n=1}^{N} (x_n - \mu_{ML})^2$$
Bayes’ Rule in Visual Perception

Examples courtesy of Jonathan Pillow, UT Austin Psychology

\[
P(M \mid D) = \frac{P(D \mid M) P(M)}{P(D)}
\]

conditional probability
“probability of M given that D occurred”

Formula for computing:
\[P(\text{what’s in the world} \mid \text{sensory data})\]
(This is what our brain wants to know!)

from
\[P(\text{sensory data} \mid \text{what’s in the world}) = \text{“likelihood”}\]

&
\[P(\text{what’s in the world}) = \text{“prior”}\]

(given by laws of physics; ambiguous because many world states could give rise to same sense data)

(given by past experience)
Same light hits the eye from both patches
Comparison patch

Same light hits the eye from both patches
Which dimples are popping out and which popping in?
Which dimples are popping out and which popping in?
Many different 3D worlds can give rise to the same 2D retinal image

The Ames Room

How does our brain go about deciding which interpretation?

\[ P(\text{image} \mid M_A) \text{ and } P(\text{image} \mid M_B) \text{ are equal! (both } M_A \text{ and } M_B \text{ could have generated this image)} \]

Let’s use Bayes’ rule:

\[ P(M_A \mid \text{image}) = P(\text{image} \mid M_A) P(M_A) \]
\[ P(M_B \mid \text{image}) = P(\text{image} \mid M_B) P(M_B) \]

Which of these is greater?
Curve Fitting Re-visited

\[ y(x_0, w) \]

\[ p(t|x_0, w, \beta) = \mathcal{N}(t|y(x_0, w), \beta^{-1}) \]
Maximum Likelihood

\[ p(t|x, w, \beta) = \prod_{n=1}^{N} \mathcal{N}(t_n | y(x_n, w), \beta^{-1}) \]

\[
\ln p(t|x, w, \beta) = -\frac{\beta}{2} \sum_{n=1}^{N} \left\{ y(x_n, w) - t_n \right\}^2 + \frac{N}{2} \ln \beta - \frac{N}{2} \ln(2\pi) \\
\underbrace{\beta E(w)}
\]

Determine \( w_{ML} \) by minimizing sum-of-squares error, \( E(w) \).

\[
\frac{1}{\beta_{ML}} = 1 \frac{1}{N} \sum_{n=1}^{N} \left\{ y(x_n, w_{ML}) - t_n \right\}^2
\]
MAP: A Step towards Bayes

\[ p(w|x, t, \alpha, \beta) \propto p(t|x, w, \beta)p(w|\alpha) \]

\[ p(w|\alpha) = \mathcal{N}(w|0, \alpha^{-1}I) = \left(\frac{\alpha}{2\pi}\right)^{(M+1)/2} \exp \left\{ -\frac{\alpha}{2} w^T w \right\} \]

\[ \beta \tilde{E}(w) = \frac{\beta}{2} \sum_{n=1}^{N} \{y(x_n, w) - t_n\}^2 + \frac{\alpha}{2} w^T w \]

Determine \( w_{MAP} \) by minimizing regularized sum-of-squares error, \( \tilde{E}(w) \).
Predictive Distribution

\[ p(t|x, \mathbf{w}_{\text{ML}}, \beta_{\text{ML}}) = \mathcal{N}(t|y(x, \mathbf{w}_{\text{ML}}), \beta_{\text{ML}}^{-1}) \]
9\textsuperscript{th} Order Polynomial

\[ M = 9 \]
Root-Mean-Square (RMS) Error: $E_{\text{RMS}} = \sqrt{\frac{2E(w^*)}{N}}$
# Polynomial Coefficients

<table>
<thead>
<tr>
<th></th>
<th>$M = 0$</th>
<th>$M = 1$</th>
<th>$M = 3$</th>
<th>$M = 9$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$w_0^*$</td>
<td>0.19</td>
<td>0.82</td>
<td>0.31</td>
<td>0.35</td>
</tr>
<tr>
<td>$w_1^*$</td>
<td></td>
<td>-1.27</td>
<td>7.99</td>
<td>232.37</td>
</tr>
<tr>
<td>$w_2^*$</td>
<td></td>
<td></td>
<td>-25.43</td>
<td>-5321.83</td>
</tr>
<tr>
<td>$w_3^*$</td>
<td></td>
<td></td>
<td>17.37</td>
<td>48568.31</td>
</tr>
<tr>
<td>$w_4^*$</td>
<td></td>
<td></td>
<td></td>
<td>-231639.30</td>
</tr>
<tr>
<td>$w_5^*$</td>
<td></td>
<td></td>
<td></td>
<td>640042.26</td>
</tr>
<tr>
<td>$w_6^*$</td>
<td></td>
<td></td>
<td></td>
<td>-1061800.52</td>
</tr>
<tr>
<td>$w_7^*$</td>
<td></td>
<td></td>
<td></td>
<td>1042400.18</td>
</tr>
<tr>
<td>$w_8^*$</td>
<td></td>
<td></td>
<td></td>
<td>-557682.99</td>
</tr>
<tr>
<td>$w_9^*$</td>
<td></td>
<td></td>
<td></td>
<td>125201.43</td>
</tr>
</tbody>
</table>
Regularization

- Penalize large coefficient values

\[
\tilde{E}(w) = \frac{1}{2} \sum_{n=1}^{N} \{y(x_n, w) - t_n\}^2 + \frac{\lambda}{2} \|w\|^2
\]
Regularization: $\ln \lambda$ vs. $E_{\text{RMS}}$