Contents

Chapter 1

Introduction

1.1 Curve Fitting

Problem 1

This can be solved by substituting the definition of:

$$y(x, \mathbf{w}) = \sum_{j=0}^{M} w_j x^j$$

into the error function and then taking the derivative.

$$\begin{aligned}
E(\mathbf{w}) &= \frac{1}{2} \sum_{n=1}^{N} (y(x, \mathbf{w}) - t_n)^2 \\
&= \frac{1}{2} \sum_{n=1}^{N} \left(\sum_{j=0}^{M} w_j x^j - t_n \right)^2 && \text{Substitute} \\
\frac{dE(\mathbf{w})}{dw_i} &= \sum_{n=1}^{N} \left(\left(\sum_{j=0}^{M} w_j x^j - t_n \right) x^i \right) && \text{Take the derivative} \\
0 &= \sum_{n=1}^{N} \left(\left(\sum_{j=0}^{M} w_j x^j - t_n \right) x^i \right) && \text{Set derivative to 0} \\
0 &= \sum_{n=1}^{N} \left(\sum_{j=0}^{M} w_j x^j x^i - t_n x^i \right) && \text{Set derivative to 0} \\
\sum_{n=1}^{N} t_n x^i &= \sum_{n=1}^{N} \sum_{j=0}^{M} w_j x^j x^i \\
\sum_{n=1}^{N} t_n x^i &= \sum_{n=1}^{N} \sum_{j=0}^{M} w_j x^{i+j}
\end{aligned}$$

Problem 2

This is solved in almost the same way we just have one additional term for the regularization so:

$$
\begin{aligned}
E(\mathbf{w}) &= \frac{1}{2}\sum_{n=1}^{N}(y(x,\mathbf{w})-t_n)^2 + \frac{\lambda}{2}\|\mathbf{w}\|^2 \\
&= \frac{1}{2}\sum_{n=1}^{N}\left(\sum_{j=0}^{M} w_j x^j - t_n\right)^2 + \frac{\lambda}{2}\|\mathbf{w}\|^2
\end{aligned}
$$

$$
\frac{dE(\mathbf{w})}{dw_i} = \sum_{n=1}^{N}\left(\left(\sum_{j=0}^{M} w_j x^j - t_n\right)x^i\right) + \lambda w_i
$$

$$
\begin{aligned}
-\lambda w_i &= \sum_{n=1}^{N}\left(\left(\sum_{j=0}^{M} w_j x^j - t_n\right)x^i\right) + \lambda w_i \\
-\lambda w_i &= \sum_{n=1}^{N}\left(\sum_{j=0}^{M} w_j x^j x^i - t_n x^i\right) \\
\sum_{n=1}^{N} t_n x^i - \lambda w_i &= \sum_{n=1}^{N}\sum_{j=0}^{M} w_j x^j x^i \\
\sum_{n=1}^{N} t_n x^i - \lambda w_i &= \sum_{n=1}^{N}\sum_{j=0}^{M} w_j x^{i+j}
\end{aligned}
$$

1.2 Probability Theory

Problem 3 The Probability of Selecting an Apple can be decomposed as:

$$
\begin{aligned}
\mathcal{P}(apple) &= \mathcal{P}(apple, red) + \mathcal{P}(apple, blue) + prob(apple, green) \\
&= \mathcal{P}(apple|red)\mathcal{P}(red) + \mathcal{P}(apple|blue)\mathcal{P}(blue) + \mathcal{P}(apple|green)\mathcal{P} \\
&= (.3)(.2) + (.5)(.2) + (.3)(.6) \\
&= .34
\end{aligned}
$$

The probability that observing an orange came from the green box can be solved using Bayes rule:

$$\begin{aligned}
\mathcal{P}(green|orange) &= \frac{\mathcal{P}(orange|green)\mathcal{P}(green)}{\mathcal{P}(orange)} \\
&= \frac{(.3)(.6)}{.66} \\
&= .27
\end{aligned}$$

Problem 5

$$\begin{aligned}
\text{Var}[X] &= \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] \\
&= \mathbb{E}\left[X^2 - 2X\mathbb{E}[X] + \mathbb{E}[X]^2\right] && \text{Distributive Law} \\
&= \mathbb{E}\left[X^2\right] + \mathbb{E}[-2X\mathbb{E}[X]] + \mathbb{E}\left[\mathbb{E}[X]^2\right] && \text{Linearity of } \mathbb{E}[X] \\
&= \mathbb{E}\left[X^2\right] + -2\mathbb{E}[X\mathbb{E}[X]] + \mathbb{E}[X]^2 && \mathbb{E}[\alpha X] = \alpha\mathbb{E}[X] \\
&= \mathbb{E}\left[X^2\right] + -2\mathbb{E}[X]^2 + \mathbb{E}[X]^2 && \mathbb{E}[X] \text{ is just another constan} \\
&= \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2
\end{aligned}$$

Problem 6

$\text{Cov}[X,Y] = \mathbb{E}[X,Y] - \mathbb{E}[X]\mathbb{E}[Y]$ But because
$X \perp Y \Rightarrow \mathbb{E}[X,Y] = \mathbb{E}[X]\mathbb{E}[Y] \Rightarrow \text{Cov}[X,Y] = 0$

Problem 7

$$
\begin{aligned}
I^2 &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\sigma^2}x^2 - \frac{1}{2\sigma^2}y^2\right) dxdy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\sigma^2}(x^2+y^2)\right) dxdy
\end{aligned}
$$

$$
\begin{aligned}
&= \int_{0}^{\infty}\int_{0}^{2\pi} \exp\left(-\frac{1}{2\sigma^2}\left[r^2\cos^2(\theta) + r^2\sin^2(\theta)\right]\right) rd\theta dr \\
&= \int_{0}^{\infty}\int_{0}^{2\pi} \exp\left(-\frac{r^2}{2\sigma^2}\right) rd\theta dr \\
&= 2\pi\int_{0}^{\infty} \exp\left(-\frac{r^2}{2\sigma^2}\right) rdr \\
&= 2\pi\int_{0}^{\infty} \exp\left(-\frac{u}{2\sigma^2}\right) \frac{1}{2}du \\
&= -2\pi\sigma^2 \exp\left(-\frac{u}{2\sigma^2}\right) |_0^\infty \\
&= 2\pi\sigma^2
\end{aligned}
$$

Now we just need to show that this normalizes the Gaussian. Take $y = x - \mu$ then

$$
\begin{aligned}
\mathcal{N}(x|\mu,\sigma^2) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) dx \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}y^2\right) dy
\end{aligned}
$$

Problem 8

$$
\begin{aligned}
\mathbb{E}[x] &= \int_{-\infty}^{\infty} \mathcal{N}(x|\mu,\sigma) x dx \\
&= \int_{-\infty}^{\infty} \frac{x}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) dx \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} (y+\mu) \exp\left(-\frac{1}{2\sigma^2}y^2\right) dy \qquad y = x-\mu
\end{aligned}
$$

Now split this into the sum of two integrals. The second integral has the form:

$$
\frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \mu \exp\left(-\frac{1}{2\sigma^2}y^2\right) dy
$$

Which is just a normalized Gaussian times μ so this is just μ. Now since we are trying to prove that his equals μ I'm fairly confident that the left integral will go to zero. Let's try and prove this.

$$
\frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} y \exp\left(-\frac{1}{2\sigma^2}y^2\right) dy
$$

Well if we visualize this function it looks odd which would imply that the integral from $[-\infty, \infty]$ is 0. Let's prove that it is odd quickly. We want to show that $f(x) = -f(-x)$ where $f(x) = y\exp\left(-\frac{1}{2\sigma^2}y^2\right)$

$$
-(-y)\exp\left(-\frac{1}{2\sigma^2}(-y)^2\right) = y\exp\left(-\frac{1}{2\sigma^2}y^2\right)
$$

Thus or integral in question is odd and evaluates to 0 yielding μ as the desired result.

Next we need to show that:

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

To do this differentiate both sides by σ^2 as follows:

$$\begin{aligned}
\frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) dx &= 1 \\
\int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) dx &= \sqrt{2\pi\sigma^2} \\
\int_{-\infty}^{\infty} \frac{d}{d\sigma^2} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) dx &= \frac{d}{d\sigma^2}\sqrt{2\pi\sigma^2} \\
\frac{1}{2\sigma^4} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) (x-\mu)^2 dx &= \frac{\pi}{\sqrt{(2\pi\sigma^2)}} \\
\int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) (x-\mu)^2 dx &= \frac{2\pi\sigma^4}{\sqrt{(2\pi\sigma^2)}} \\
\frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) (x-\mu)^2 dx &= \sigma^2 \\
\mathrm{Var}\left[x\right] &= \sigma^2
\end{aligned}$$

The left hand side of this equation is just the definition of variance. Now to complete the proof of (1.50) just use the alternate formulation of variance:

$$\begin{aligned}
\mathbb{E}\left[(x-\mu)^2\right] &= \mathbb{E}\left[x^2\right] - \mathbb{E}\left[x\right]^2 \\
&= \mathbb{E}\left[x^2\right] - \mu^2 && \text{By part 1} \\
\sigma^2 &= \mathbb{E}\left[x^2\right] - \mu^2 && \text{By above} \\
\mathbb{E}\left[x^2\right] &= \sigma^2 + \mu^2
\end{aligned}$$

Problem 9

$$\begin{aligned}
\mathcal{N}(x|\mu,\sigma^2) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) \\
\frac{d\mathcal{N}(x|\mu,\sigma^2)}{dx} &= -\frac{1}{\sigma^2}\mathcal{N}(x|\mu,\sigma^2)(x-\mu) \\
0 &= \mathcal{N}(x|\mu,\sigma^2)(x-\mu)
\end{aligned}$$

But since $\mathcal{N}(x|\mu,\sigma^2)(x-\mu) > 0$ the only term that matters is $(x-\mu)$, which goes to 0 at $x = \mu$

The proof for the multivariate case is almost identical and is omitted

Problem 10

$$\begin{aligned}
\log p(\mathbf{x}|\mu,\sigma^2) &= -\frac{1}{\sqrt{2\pi\sigma^2}}\sum_{n=1}^{N}(x_n-\mu)^2 - \frac{N}{2}\log\sigma^2 - \frac{N}{2}\log(2\pi) \\
\frac{d\log p(\mathbf{x}|\mu,\sigma^2)}{d\mu} &= -\frac{1}{\sqrt{\pi\sigma^2}}\sum_{n=1}^{N}(x_n-\mu) \\
0 &= \sum_{n=1}^{N}(x_n-\mu) \\
&= \sum_{n=1}^{N}x_n - \sum_{n=1}^{N}\mu \\
\mu &= \sum_{n=1}^{N}x_n
\end{aligned}$$

Problem 11