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.

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

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

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

Problem 6

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

Problem 7

Problem 8

Problem 9

Problem 10