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

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

Problem 6

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

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

Just take the derivative with respect to μ of our function:

$$\ln(x|\mu,\sigma^2) = -\frac{1}{2\sigma^2}\sum_{n=1}^{N}(x_n-\mu)^2 - \frac{N}{2}\ln\sigma^2 - \frac{N}{2}\ln(2\pi)$$

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

Now repeat the same for σ^2

$$\frac{d}{d\mu}\ln(x|\mu,\sigma^2) = \frac{1}{2\sigma^4}\sum_{n=1}^{N}(x_n-\mu)^2 - \frac{N}{2\sigma^2}$$
$$0 = \frac{1}{2\sigma^4}\sum_{n=1}^{N}(x_n-\mu)^2 - \frac{N}{2\sigma^2}$$
$$\frac{N}{2\sigma^2} = \frac{1}{2\sigma^4}\sum_{n=1}^{N}(x_n-\mu)^2$$
$$\sigma^2 = \frac{1}{N}\sum_{n=1}^{N}(x_n-\mu)^2$$

1.3 Curse of Dimensionality

Problem 17

The gamma function is defined as:

$$\Gamma(x) = \int_0^\infty u^{x-1}e^{-u}du$$

We use integration by parts with the following substitutions

$$u = u^{x-1} \qquad v = -e^{-u}$$
$$du = (x-1)u^{x-2}du \qquad dv = e^{-u}du$$

By integration by parts we have:

$$\begin{aligned}
\Gamma(x) &= -u^{x-1}e^{-u}|_0^\infty + (x-1)\int_0^\infty u^{x-2}e^{-u}du \\
&= (x-1)\int_0^\infty u^{x-2}e^{-u}du \\
&= (x-1)\Gamma(x-1) \\
&= z\Gamma(z) && z = x+1
\end{aligned}$$

Now see that $\Gamma(1) = 1$

$$\begin{aligned}
\Gamma(1) &= \int_0^\infty u^0 e^{-u}du \\
&= \int_0^\infty e^{-u}du \\
&= -e^{-u}|_0^\infty \\
&= 1
\end{aligned}$$

To show that for all $x \in \mathbb{Z}^+$ $\Gamma(x+1) = x!$ we use an inductive argument. Our base case is $\Gamma(1) = 1$ which we have already done. Now let's assume that our inductive hypothesis is true and $\Gamma(x+1) = x!$. Then $\Gamma(x+2) = (x+1)\Gamma(x+1) = (x+1)x!$ and we are done.

Problem 18

We want to solve for S_D the surface area of a D dimensional sphere.

$$\prod_{i=1}^{D}\int_{-\infty}^{\infty} e^{-x_i^2} = S_D\int_0^\infty e^{-r^2}r^{D-1}dr$$

Let's solve the left hand side first. From problem 7 above we know that

$$\int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}x^2} = \sqrt{2\pi\sigma^2}$$

By setting $\sigma^2 = \frac{1}{2}$ we get that

$$\begin{aligned}\prod_{i=1}^{D} \int_{-\infty}^{\infty} e^{-x^2} &= \prod_{i=1}^{D} \sqrt{\pi} \\ &= \pi^{\frac{D}{2}}\end{aligned}$$

Now let's look at the right hand side of the equation. We can tease this function into the Gamma function by making the substitution $u = r^2$ which implies that $dr = \frac{1}{2}u^{-1/2}$. With this insight we see that:

$$\begin{aligned}\int_{0}^{\infty} e^{-r^2} r^{D-1} dr &= \frac{1}{2}\int_{0}^{\infty} e^{-u} u^{\frac{D}{2}-1} du \\ &= \frac{1}{2}\Gamma(D/2)\end{aligned}$$

Now just use simple algebra:

$$\begin{aligned}\pi^{\frac{D}{2}} &= S_D \frac{1}{2}\Gamma(D/2) \\ S_D &= \frac{2\pi^{D/2}}{\Gamma(D/2)}\end{aligned}$$

.

Next integrate the surface area with respect to the radius in order to get the equation for volume:

$$
\begin{aligned}
V_D &= S_D \int_0^1 r^{D-1} dr \\
&= S_D \left.\frac{r}{D}\right|_0^1 \\
&= \frac{S_D}{D}
\end{aligned}
$$

Now to derive some common known volumes. For D =2:

$$
\begin{aligned}
\frac{2\pi^{D/2}}{D\Gamma(D/2)} &= \frac{2\pi}{2\Gamma(1)} \\
&= \pi
\end{aligned}
$$

For D=3:

$$
\begin{aligned}
\frac{2\pi^{D/2}}{D\Gamma(D/2)} &= \frac{2\pi^{3/2}}{3\Gamma(3/2)} \\
&= \frac{2\pi^{3/2}}{3\frac{\sqrt{\pi}}{2}} \\
&= \frac{4\pi}{3}
\end{aligned}
$$

Problem 19

From problem 18 we know what the volume of a hypersphere is. The volume of a D-dimensional cube is simply l^D where l is the length of a side. In our case this is $2a$ so the ratio is:

$$\frac{\text{Volume Sphere}}{\text{Volume Cube}} = \frac{2\pi^{D/2}a^D}{(2a)^D D\Gamma(D/2)} = \frac{\pi^{D/2}}{2^{D-1}D\Gamma(D/2)}$$

To show that the ratio goes to zero I'm going to eschew Stirling's formula because there is a much easier way to show this.... Begin by noticing that the above equation is always positive. Therefore we know that $0 \leq \frac{\pi^{D/2}}{2^{D-1}D\Gamma(D/2)}$ $\forall D \in \mathbb{Z}$. Now let's replace the gamma function by the factorial function because they are identical for integer operands. I will also drop the D and 2^{D-1} from the bottom which increases the total value of the expression giving me:

$$0 \leq \frac{\pi^{D/2}}{2^{D-1}D\Gamma(D/2)} \leq \frac{\pi^{D/2}}{(D/2)!}$$
$$0 \leq \frac{\pi^x}{x2^{2x}\Gamma(x)} \leq \frac{\pi^x}{x!}$$

Now we just need to show that $\lim_{x\to\infty} \frac{\pi^x}{x!} = 0$. Which is easy! Notice that:

$$0 < \frac{\pi^x}{x!}$$
$$\frac{\pi^x}{x!} = \frac{\pi}{1}\cdot\frac{\pi}{2}\cdot\frac{\pi}{3}\cdot\ldots\cdot\frac{\pi}{x} < \frac{\pi}{1}\cdot\frac{\pi}{2}\cdot\frac{\pi}{3}\cdot\frac{\pi}{4}\cdot\frac{\pi}{4}\cdot\ldots\cdot\frac{\pi}{4} = \frac{\pi}{6}\left(\frac{\pi}{4}\right)^{n-3}$$

Now we know that:

$$\lim_{x\to\infty} \left(\frac{\pi}{4}\right)^{n-3} = 0$$

Therefore by the squeeze theorem so does $\frac{\pi^{D/2}}{2^{D-1}D\Gamma(D/2)}$ and we are done.

To show that the ratio of the distance from the center to a corner to the distance from the center to a side is $\sqrt{D}$ start by centering the cube at the origin. Then we know that one corner will be at $(a, a, ..., a)$ and the distance to that corner will be $a\sqrt{D}$ in euclidean space. Now look at the distance to a side. A side will be at the point $(a, 0, 0, ...)$ which will be distance a from the origin in Euclidean space so the ratio is $\sqrt{D}$

Problem 20

1.4 Decision Theory

Problem 21

First we prove that given two non negative numbers a, b and $a \leq b$ $\Rightarrow a \leq \sqrt{ab}$:

$$\begin{aligned} a &\leq ab \\ a^2 &\leq ab \\ |a| &\leq \sqrt{ab} \\ a &\leq \sqrt{ab} \end{aligned}$$

the probability of a mistake is the probability of making an error over all classification sub regions.

$$p(mistake) = \int_{R_1} p(x, C_1)dx + \int_{R_2} p(x, C_2)dx$$

We know that over R_1 $p(C_1|x) \geq p(C_2|x)$. By our first proof we know that:

$$\begin{aligned} p(C_2|x) &\leq \sqrt{p(C_1|x)p(C_2|x)} \\ p(C_2|x)p(x) &\leq p(x)\sqrt{p(C_1|x)p(C_2|x)} \\ p(x, C_2) &\leq \sqrt{p(C_1|x)p(C_2|x)p(x)^2} \\ &\leq \sqrt{p(x, C_1)p(x, C_2)} \end{aligned}$$

By an identical argument over R_2 $p(C_1|x) \leq p(C_2|x)$ and

$$p(x, C_1) \leq \sqrt{p(C_1|x)p(C_2|x)p(x)^2}$$

substituting these into the definition of a mistake probability:

$$\begin{aligned} p(mistake) &\leq \int_{R_1} \sqrt{p(x, C_1)p(x, C_2)}dx + \int_{R_2} \sqrt{p(x, C_1)p(x, C_2)}dx \\ &\leq \int \sqrt{p(x, C_1)p(x, C_2)}dx \end{aligned}$$

Problem 22

The loss matrix $L_{kj} = 1 - I_{kj}$ can be visualized as:

$$L = \begin{bmatrix} 0 & 1 & 1 & \cdots & 1 \\ 1 & 0 & 1 & \cdots & 1 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 1 \\ 1 & \cdots & \cdots & 1 & 0 \end{bmatrix}$$

This is interpreted as there is no penalty for choosing the correct class, however any type of misclassification is penalized the same. This means that we just want to choose the class with the largest posterior probability because it will be the most likely to be correct, if it is incorrect it is no more incorrect than any other non correct guess. To see this formally we use the delta function which is defined as:

$$\delta_{kj} = \begin{cases} 0, & \text{for } k = j \\ 1, & \text{for} \;\; \text{else} \end{cases}$$

Using this notation we can write $L_{kj} = 1 - \delta_{kj}$ This makes the function we want to minimize:

$$\begin{aligned}
\sum_k L_{kj} p(C_k|x) &= \sum_k (1-\delta_{kj}) p(C_k|x) \\
&= \sum_k p(C_k|x) - \delta_{kj} p(C_k|x) \\
&= \sum_k p(C_k|x) - \sum_k \delta_{kj} p(C_k|x) \\
&= 1 - \sum_k \delta_{kj} p(C_k|x)
\end{aligned}$$

In order to minimize this function we want $p(C_k|x)$ to be as large as possible so we choose the class which has the largest posterior probability.

Problem 23