Chapter 1 - Introduction

Kristian Wichmann

August 14, 2017

1 Exercise 1.1

Suppose we are given a training set, of N observation points:

$$\mathbf{x} = (x_1, \cdots, x_n)^t, \quad \mathbf{t} = (t_1, \cdots, t_n)^t \tag{1.1}$$

We think of the t's as being dependent on the x'es: $t = t(x)$. We wish to devise a model through polynomial curve fitting of order M. I.e. our model is of the form:

$$y(x, \mathbf{w}) = w_0 + w_1 x + w_2 x^2 + \cdots + w_M x^M = \sum_{j=0}^{M} w_j x^j \tag{1.2}$$

Here the parameters $\mathbf{w} = (w_0, \cdots, w_M)^t$ are known as the weights. We wish to find the model which minimizes the following error function:

$$E(\mathbf{w}) = \frac{1}{2} \sum_{n=1}^{N} [y(x_n, \mathbf{w}) - t_n)]^2 \tag{1.3}$$

1.1 Solution

To minimize the error function, we differentiate with respect to the i'th weight, using the chain rule:

$$\frac{\partial E}{\partial w_i} = \frac{1}{2} \sum_{n=1}^{N} \frac{\partial}{\partial w_i} [y(x_n, \mathbf{w}) - t_n)]^2 = \sum_{n=1}^{N} [y(x_n, \mathbf{w}) - t_n)] \frac{\partial y(x_n, \mathbf{w})}{\partial w_i} \tag{1.4}$$

The last derivative is:

$$\frac{\partial y(x_n, \mathbf{w})}{\partial w_i} = \frac{\partial}{\partial w_i} \sum_{j=0}^{M} w_j x_n^j = \sum_{j=0}^{M} \delta_{ij} x_n^j = x_n^i \tag{1.5}$$

Now equation 1.4 becomes:

$$\frac{\partial E}{\partial w_i} = \sum_{n=1}^{N} \left[\sum_{j=0}^{M} w_j x_n^j - t_n \right] x_n^i = \sum_{n=1}^{N} \left[\sum_{j=0}^{M} w_j x_n^{i+j} - t_n x_n^i \right] \tag{1.6}$$

Now define:

$$A_{ij} = \sum_{n=1}^{N} x_n^{i+j} \quad T_i = \sum_{n=1}^{N} t_n x_n^i \tag{1.7}$$

Then we can rewrite:

$$\frac{\partial E}{\partial w_i} = \sum_{j=0}^{M} A_{ij} w_j - T_i \tag{1.8}$$

Setting this derivative equal to zero, we get:

$$\sum_{j=0}^{M} A_{ij} w_j = T_i \tag{1.9}$$

2 Exercise 1.2

Consider the same problem as last exercise, but with an additional regularization term added to the error function:

$$\tilde{E}(\mathbf{w}) = \frac{1}{2} \sum_{n=1}^{N} [y(x_n, \mathbf{w}) - t_n)]^2 + \frac{\lambda}{2} ||\mathbf{w}||^2 \tag{2.1}$$

We wish to find the equations the weights must satisfy in this case.

2.1 Solution

The derivative of the error function is modified:

$$\frac{\partial \tilde{E}}{\partial w_i} = \frac{\partial E}{\partial w_i} + \frac{\partial}{\partial w_i} \left[\frac{\lambda}{2} ||\mathbf{w}||^2 \right] \tag{2.2}$$

The last correction term is:

$$\frac{\lambda}{2} \frac{\partial}{\partial w_i} (w_0^2 + w_1^2 + \cdots + w_M^2) = \lambda w_i \tag{2.3}$$

According to equation 1.8 we have:

$$\frac{\partial \tilde{E}}{\partial w_i} = \sum_{j=0}^{M} A_{ij} w_j - T_i + w_i \tag{2.4}$$

Table 1: Content of the boxes (Exercise 1.3)

	r	b	g
Apples	3	1	3
Oranges	4	1	3
Limes	3	0	4

Setting this equal to zero we get:

$$\sum_{j=0}^{M} A_{ij} w_j + \lambda w_i = T_i \tag{2.5}$$

We may absorb the λ term into the A matrix by the following modification:

$$\tilde{A}_{ij} = A_{ij} + \lambda \delta_{ij} \tag{2.6}$$

Then we get:

$$\sum_{j=0}^{M} \tilde{A}_{ij} w_j = T_i \tag{2.7}$$

3 Exercise 1.3

We have three colored boxes, r (red), b (blue), and g (green). The contents of the three boxes are shown in table 1. A box is chosen at random according to the following probability distribution:

$$p(r) = \frac{1}{5}, \quad p(b) = \frac{1}{5}, \quad p(g) = \frac{3}{5} \tag{3.1}$$

Next, a random piece of fruit (equal chance for each piece) is chosen from the box.

1. What is the probability of the fruit being an apple?
2. Given that the selected fruit is an orange, what is that probability that the green box was picked?

3.1 Solution 1

The probability of getting an apple (A) is:

$$p(A) = p(A|r)p(r) + p(A|b)p(b) + p(A|g)p(g) \tag{3.2}$$

Inserting:

$$p(A) = \frac{3}{10}\cdot\frac{1}{5} + \frac{1}{2}\cdot\frac{1}{5} + \frac{3}{10}\cdot\frac{3}{5} = \frac{3}{50} + \frac{1}{10} + \frac{9}{50} = \frac{3+5+9}{50} = \frac{17}{50} = 34\% \tag{3.3}$$

3.2 Solution 2

We wish to find $p(g|O)$, where 'O' is short for orange. For this we need Bayes' rule:

$$p(g|O) = \frac{p(O|g)p(g)}{p(O)} \tag{3.4}$$

The denominator can be found similarly to problem 1:

$$p(O) = p(O|r)p(r) + p(O|b)p(b) + p(O|g)p(g) \tag{3.5}$$

Inserting:

$$p(O) = \frac{4}{10} \cdot \frac{1}{5} + \frac{1}{2} \cdot \frac{1}{5} + \frac{3}{10} \cdot \frac{3}{5} = \frac{4+5+9}{50} = \frac{18}{50} = 36\% \tag{3.6}$$

Now we have:

$$p(g|O) = \frac{3/10 \cdot 3/5}{18/50} = \frac{9 \cdot 50}{50 \cdot 18} = \frac{9}{18} = \frac{1}{2} \tag{3.7}$$

4 Exercise 1.4

Let $p_x(x)$ be a probability density function for a continuous variable x. Consider a variable transformation $x = g(y)$. The pdf for the transformed variable is then:

$$p_y(y) = p_x(x) \left| \frac{dx}{dy} \right| = p_x(g(y))|g'(y)| \tag{4.1}$$

Show that in general, a non-linear transformation will change the location of a maximum of the pdf, while a linear one will not.

4.1 Solution

If p_x has a maximum at $x = \hat{x}$ then we must have:

$$\frac{dp_x(\hat{x})}{dx} = 0 \tag{4.2}$$

Now, consider the derivative of p_y:

$$\frac{dp_y(y)}{dy} = \frac{dp_x}{dx}\frac{dg}{dy}|g'(y)| + p_x(g(y))\mathrm{sgn}(g'(y))g''(y) \tag{4.3}$$

If $y = g(\hat{x})$, the first term is zero, but the second need not be. Therefore, the location of a maximum will generally move. However, for a linear transformation the second derivative is zero, and so the second term becomes zero as well.

5 Exercise 1.5

The variance of a continuous random variable $f(x)$ is defined as:

$$\mathrm{var}[f(x)] = \mathbb{E}[(f(x) - \mathbb{E}[f(x)])^2] \tag{5.1}$$

Show that this may be written:

$$\mathrm{var}[f(x)] = \mathbb{E}[f(x)^2] - \mathbb{E}[f(x)]^2 \tag{5.2}$$

5.1 Solution

Expand the square inside the expectation value:

$$(f(x) - \mathbb{E}[f(x)])^2 = f(x)^2 + \mathbb{E}[f(x)]^2 - 2f(x)\mathbb{E}[f(x)] \tag{5.3}$$

The expectation values is:

$$\mathbb{E}[f(x)^2] + \mathbb{E}[f(x)]^2 - 2\mathbb{E}[f(x)] \cdot \mathbb{E}[f(x)] = \mathbb{E}[f(x)^2] - \mathbb{E}[f(x)]^2 \tag{5.4}$$

6 Exercise 1.6

Show that two independent random variable x and y have zero covariance.

6.1 Solution

The covariance can be written:

$$\int\int (x - \mathbb{E}[x])(y - \mathbb{E}[y])p(x, y)dx\ dy \tag{6.1}$$

Here $p(x, y)$ is the joint distribution function of x and y. But since these are independent $p(x, y) = p(x)p(y)$. So this integral splits into two parts:

$$\int\int (x-\mathbb{E}[x])(y-\mathbb{E}[y])p(x)p(y)dx\ dy = \int (x-\mathbb{E}[x])dx \cdot \int (y-\mathbb{E}[y])dy \tag{6.2}$$

Each of these factors is zero, and therefore so is the covariance.

7 Exercise 1.7

Show that:

$$I = \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\sigma^2}x^2\right) dx = \sqrt{2\pi\sigma^2} \tag{7.1}$$

7.1 Solution

Consider the square of I:

$$I^2 = \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\sigma^2}x^2\right) dx \cdot \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\sigma^2}y^2\right) dy \tag{7.2}$$

Rearrange the order:

$$I^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\sigma^2}x^2\right) \exp\left(-\frac{1}{2\sigma^2}y^2\right) dx\ dy = \tag{7.3}$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\sigma^2}(x^2+y^2)\right) dx\ dy \tag{7.4}$$

The may be regarded as an integral over the plane. We may now shift to polar coordinates, noting that $x^2 + y^2 = r^2$ and $dx\ dy = r\ dr\ d\theta$:

$$I^2 = \int_0^{2\pi}\int_0^{\infty} \exp\left(-\frac{r^2}{2\sigma^2}\right) r\ dr\ d\theta = 2\pi \int_0^{\infty} \exp\left(-\frac{r^2}{2\sigma^2}\right) r\ dr \tag{7.5}$$

For the r-integral, substitute $u = r^2$. Then $du/dr = 2r$, and so $dr = \frac{1}{2r}du$:

$$\int_0^{\infty} \exp\left(-\frac{r^2}{2\sigma^2}\right) r\ dr = \int_0^{\infty} \exp\left(-\frac{u}{2\sigma^2}\right) r\frac{1}{2r}du = \frac{1}{2}\int_0^{\infty} \exp\left(-\frac{u}{2\sigma^2}\right) du \tag{7.6}$$

This is elemental to integrate:

$$\frac{1}{2}\left[-2\sigma^2 \exp\left(-\frac{u}{2\sigma^2}\right)\right]_0^{\infty} = \sigma^2 \tag{7.7}$$

Hence $I^2 = 2\pi\sigma^2$, and therefore $I = \sqrt{2\pi\sigma^2}$.

8 Exercise 1.8

The univariate Gaussian distribution with mean μ and variance σ^2 has the pdf:

$$\mathcal{N}(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \tag{8.1}$$

1. Show that the expectation value is actually μ.
2. Show that $\mathbb{E}[x^2] = \mu^2 + \sigma^2$.
3. Show that the variance is actually σ^2.

8.1 Solution 1

We need to evaluate the integral:

$$\mathbb{E}[x] = \int_{-\infty}^{\infty} \mathcal{N}(x|\mu, \sigma^2) x \, dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) x \, dx \tag{8.2}$$

Now apply the substitution $u = x - \mu$. This means $du = dx$. Also $x = u + \mu$:

$$\mathbb{E}[x] = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \exp\left(-\frac{u^2}{2\sigma^2}\right) (u + \mu) du \tag{8.3}$$

We may split the integral into two:

$$\int_{-\infty}^{\infty} \exp\left(-\frac{u^2}{2\sigma^2}\right) u \, du + \mu \int_{-\infty}^{\infty} \exp\left(-\frac{u^2}{2\sigma^2}\right) du \tag{8.4}$$

The first is an uneven function, and so the integral is zero. The second is equal to $\mu\sqrt{2\pi\sigma^2}$ according to the last exercise. So all in all we get:

$$\mathbb{E}[x] = \mu \tag{8.5}$$

8.2 Solution 2

The normalization condition is:

$$\int_{-\infty}^{\infty} \mathcal{N}(x|\mu, \sigma^2) \, dx = 1 \tag{8.6}$$

Now differentiate on both sides of the equation with respect to σ^2. The right side is obvious zero, while the left side is:

$$\int_{-\infty}^{\infty} \frac{\partial}{\partial(\sigma^2)} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx = \tag{8.7}$$

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \frac{-1}{2(\sigma^2)^{3/2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx + \tag{8.8}$$

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \frac{-(x-\mu)^2}{2} \frac{-1}{(\sigma^2)^2} dx = \tag{8.9}$$

$$-\frac{1}{\sigma^2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx + \tag{8.10}$$

$$\frac{1}{\sigma^4} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) (x-\mu)^2 \, dx \tag{8.11}$$

The first integral is simply 1, while the second can be split further into three by expanding the square:

$$(x-\mu)^2 = x^2 + \mu^2 - 2x\mu \tag{8.12}$$

The x^2-integral is what we're trying to calculate. The mu^2-integral is a constant times 1 again. And finally the mixed term can be found from part 1. All in all:

$$\begin{aligned}
-\frac{1}{\sigma^2} + \frac{1}{\sigma^4}\left[\mathbb{E}[x^2] + \mu^2 - 2\mu\cdot\mu\right] = 0 \Leftrightarrow \quad &(8.13)\\
-1 + \frac{1}{\sigma^2}\left[\mathbb{E}[x^2] - \mu^2\right] = 0 \Leftrightarrow \quad &(8.14)\\
\frac{1}{\sigma^2}\left[\mathbb{E}[x^2] - \mu^2\right] = 1 \Leftrightarrow \quad &(8.15)\\
\mathbb{E}[x^2] - \mu^2 = \sigma^2 \Leftrightarrow \quad &(8.16)\\
\mathbb{E}[x^2] = \mu^2 + \sigma^2 \quad &(8.17)
\end{aligned}$$

8.3 Solution 3

We can now find the variance by applying the formula from the exercise above, and using the previous solutions:

$$\mathrm{var}[x] = \mathbb{E}[x^2] - \mathbb{E}[x]^2 = \mu^2 + \sigma^2 - \mu^2 = \sigma^2 \tag{8.18}$$

9 Exercise 1.9

Show that the mode of the univariate Gaussian distribution $\mathcal{N}(x|\mu,\sigma^2)$ has mode μ. Similarly, show the that multivariate Gaussian $\mathcal{N}(\mathbf{x}|\mu,\mathbf{\Sigma})$ has mode μ.

9.1 Solution

The mode is where the pdf is maximal. Differentiate with respect to x:

$$\frac{\partial}{\partial x}\mathcal{N}(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}\frac{\partial}{\partial x}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \tag{9.1}$$

Using the chain rule this is equal to:

$$\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)\left(-\frac{2}{2\sigma^2}(x-\mu)\right) \tag{9.2}$$

This may only be equal to zero if $x - \mu = 0 \Leftrightarrow x = \mu$.

For the multivariate Gaussian in D dimensions, the pdf is:

$$\mathcal{N}(\mathbf{x}|\mu, \mathbf{\Sigma}) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\mathbf{\Sigma}|} \exp\left\{ -\frac{1}{2}(\mathbf{x} - \mu)^t \mathbf{\Sigma}^{-1} (\mathbf{x} - \mu) \right\} \tag{9.3}$$

Again, the normalization constant doesn't matter, and since the exponential is monotonically increasing, this comes down to minimizing the quadratic form. Written in coordinate form, this is:

$$-\frac{1}{2}\sum_{i=1}^{D}\sum_{j=1}^{D}(\mathbf{x} - \mu)_i \mathbf{\Sigma}_{ij}^{-1} (\mathbf{x} - \mu)_j = -\frac{1}{2}\sum_{i=1}^{D}\sum_{j=1}^{D}(x_i - \mu_i)\mathbf{\Sigma}_{ij}^{-1}(x_j - \mu_j) \tag{9.4}$$

Now differentiate with respect to x_k:

$$\frac{\partial}{\partial x_k}\left[-\frac{1}{2}\sum_{i=1}^{D}\sum_{j=1}^{D}(x_i - \mu_i)\mathbf{\Sigma}_{ij}^{-1}(x_j - \mu_j) \right] = \tag{9.5}$$

$$-\frac{1}{2}\sum_{i=1}^{D}\sum_{j=1}^{D}\frac{\partial}{\partial x_k}(x_i - \mu_i)\mathbf{\Sigma}_{ij}^{-1}(x_j - \mu_j) + \tag{9.6}$$

$$-\frac{1}{2}\sum_{i=1}^{D}\sum_{j=1}^{D}(x_i - \mu_i)\mathbf{\Sigma}_{ij}^{-1}\frac{\partial}{\partial x_k}(x_j - \mu_j) = \tag{9.7}$$

$$-\frac{1}{2}\sum_{i=1}^{D}\sum_{j=1}^{D}\left[\delta_{ik}\mathbf{\Sigma}_{ij}^{-1}(x_j - \mu_j) + (x_i - \mu_i)\mathbf{\Sigma}_{ij}^{-1}\delta_{jk}\right] \tag{9.8}$$

Cancelling the deltas and renaming the first summation this is:

$$-\frac{1}{2}\sum_{i=1}^{D}\left[\mathbf{\Sigma}_{ki}^{-1}(x_i - \mu_i) + \mathbf{\Sigma}_{ik}^{-1}(x_i - \mu_i)\right] \tag{9.9}$$

Since the covariance matrix is symmetric, so is the inverse, and we get:

$$-\sum_{i=1}^{D}\mathbf{\Sigma}_{ki}^{-1}(x_i - \mu_i) \tag{9.10}$$

Back in matrix form this should be zero:

$$-\mathbf{\Sigma}^{-1}(\mathbf{x} - \mu) = 0 \tag{9.11}$$

Now multiply both sides by $\mathbf{\Sigma}$ to get that $\mathbf{x} - \mu = 0 \Leftrightarrow \mathbf{x} = \mu$.

10 Exercise 1.10

Assuming the random variables x and z to independent, show that:

1. $\mathbb{E}[x+z] = \mathbb{E}[x] + \mathbb{E}[z]$
2. $\text{var}[x+z] = \text{var}[x] + \text{var}[z]$

10.1 Solution 1

Assuming the variables to be continuous, we have:

$$\mathbb{E}[x+z] = \int\int f(x,z)(x+z)dx\ dz \tag{10.1}$$

But since the variables are independent this is:

$$\begin{aligned}
\mathbb{E}[x+z] = &\int\int f(x)f(z)(x+z)dx\ dz = &(10.2)\\
&\int\int f(x)f(z)x\ dx\ dz + \int\int f(x)f(z)z\ dx\ dz = &(10.3)\\
&\int f(z)\int f(x)x\ dx\ dz + \int f(x)\int f(z)z\ dz\ dx = &(10.4)\\
&\int f(z)\mathbb{E}[x]dz + \int f(x)\mathbb{E}[z]dx = &(10.5)\\
&\mathbb{E}[x]\int f(z)dz + \mathbb{E}[z]\int f(x)dx = \mathbb{E}[x] + \mathbb{E}[z] &(10.6)
\end{aligned}$$

Similarly for discrete variables, where the integrals are replaced by sums.

10.2 Solution 2

If x and z are independent, then so are x^n and z^m. If follows from above, that:

$$\mathbb{E}[x^n + z^m] = \mathbb{E}[x^n] + \mathbb{E}[z^m] \tag{10.7}$$

So, using the cases where $n, m = 1, 2$, the variance is:

$$\text{var}[x+z] = \mathbb{E}[(x+z)^2] - \mathbb{E}[x+z]^2 = \mathbb{E}[x^2] + \mathbb{E}[z^2] + 2\mathbb{E}[xz] - \mathbb{E}[x+z]^2 \tag{10.8}$$

But we also know that:

$$\mathbb{E}[x+z]^2 = (\mathbb{E}[x] + \mathbb{E}[z])^2 = \mathbb{E}[x]^2 + \mathbb{E}[z]^2 + 2\mathbb{E}[x]\mathbb{E}[z] \tag{10.9}$$

So:

$$\mathrm{var}[x+z] = \mathrm{var}[x] + \mathrm{var}[z] + 2(\mathbb{E}[xz] - \mathbb{E}[x]\mathbb{E}[z]) \tag{10.10}$$

So if we can show $\mathbb{E}[xz] = \mathbb{E}[x]\mathbb{E}[z]$ we're done:

$$\mathbb{E}[xz] = \int\int f(x,z)xz\ dx\ dz = \tag{10.11}$$

$$\int\int f(x)f(z)xz\ dx\ dz = \int f(x)x\ dx \int f(z)z\ dz = \mathbb{E}[x]\mathbb{E}[z] \tag{10.12}$$