Some Notes About Linear Regression, Decision Theory and Info Theory

Professor: Daniel S. Menasché

1 Linear Regression: The Simplest Case

Let $y = ax + b$.

$$e = \begin{pmatrix} e_1 \\ e_2 \end{pmatrix} = \begin{pmatrix} y_0 \\ y_1 \end{pmatrix} - b \begin{pmatrix} 1 \\ 1 \end{pmatrix} - a \begin{pmatrix} x_0 \\ x_1 \end{pmatrix} \quad (1)$$

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} y_0 - b - ax_0 \\ y_1 - b - ax_1 \end{pmatrix} \quad (2)$$

Alternatively,

$$\begin{aligned}
\begin{pmatrix} b & a \end{pmatrix} &= \begin{pmatrix} y_0 & y_1 \end{pmatrix} \begin{pmatrix} 1 & x_0 \\ 1 & x_1 \end{pmatrix} \left[\left[\begin{pmatrix} 1 & 1 \\ x_0 & x_1 \end{pmatrix} \begin{pmatrix} 1 & x_0 \\ 1 & x_1 \end{pmatrix} \right]^{-1} \right]' & (3) \\
&= \begin{pmatrix} y_0 & y_1 \end{pmatrix} \begin{pmatrix} 1 & x_0 \\ 1 & x_1 \end{pmatrix} \left[\begin{pmatrix} 2 & x_0 + x_1 \\ x_0 + x_1 & x_0^2 + x_1^2 \end{pmatrix}^{-1} \right]' & (4) \\
&= \begin{pmatrix} y_0 & y_1 \end{pmatrix} \begin{pmatrix} 1 & x_0 \\ 1 & x_1 \end{pmatrix} \frac{1}{(x_1 - x_0)^2} \begin{pmatrix} x_0^2 + x_1^2 & -x_0 - x_1 \\ -x_0 - x_1 & 2 \end{pmatrix} & (5) \\
&= \begin{pmatrix} y_0 & y_1 \end{pmatrix} \frac{1}{(x_1 - x_0)} \begin{pmatrix} +x_1 & -1 \\ -x_0 & +1 \end{pmatrix} & (6) \\
&= \begin{pmatrix} (y_0 x_1 - y_1 x_0)/(x_1 - x_0) & (y_1 - y_0)/(x_1 - x_0) \end{pmatrix} & (7)
\end{aligned}$$

2 Decision Theory

Example taken from Chapter 1 by Duda and Hart

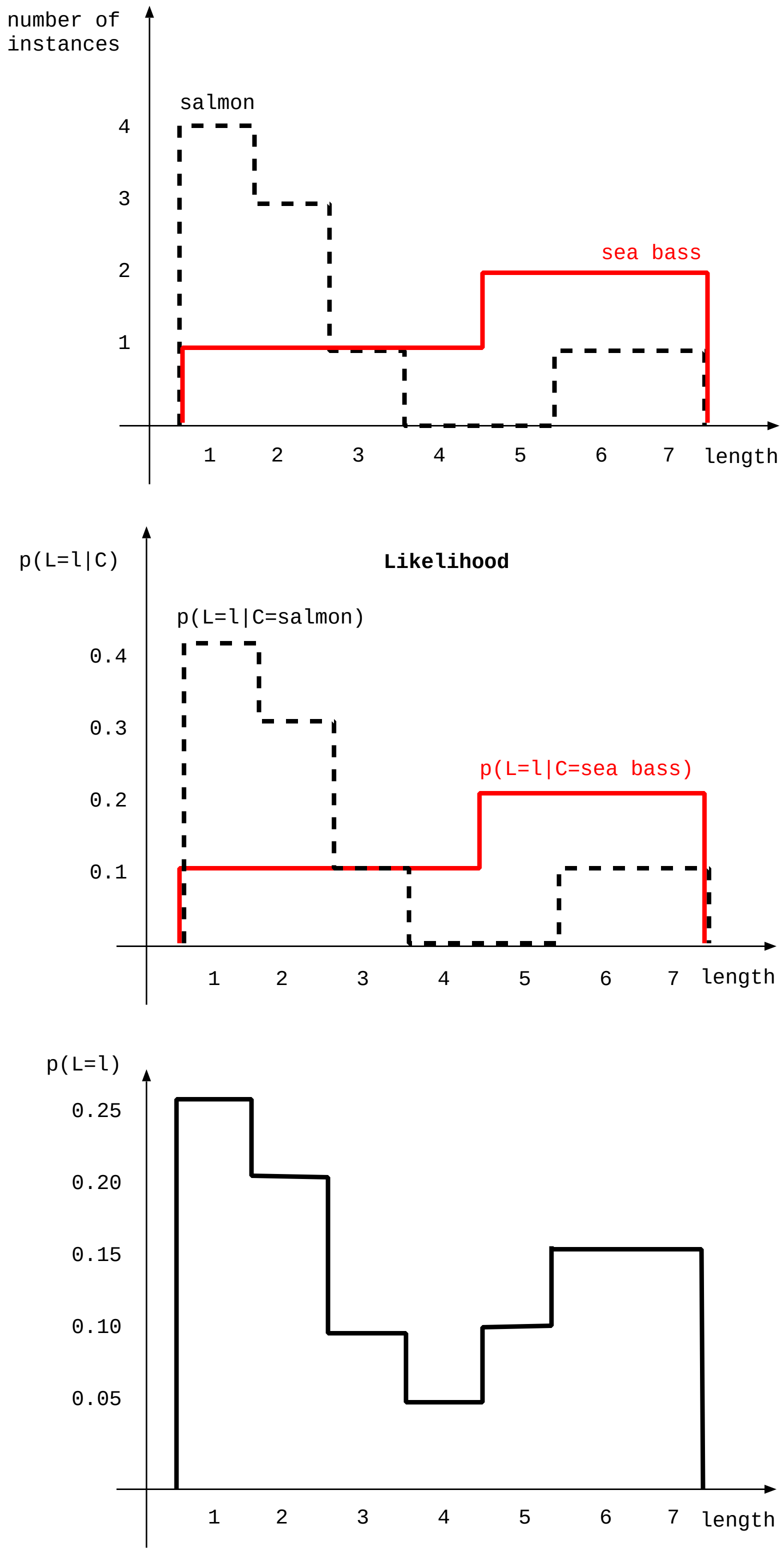
Consider the problem of classifying fish between salmon and sea bass.

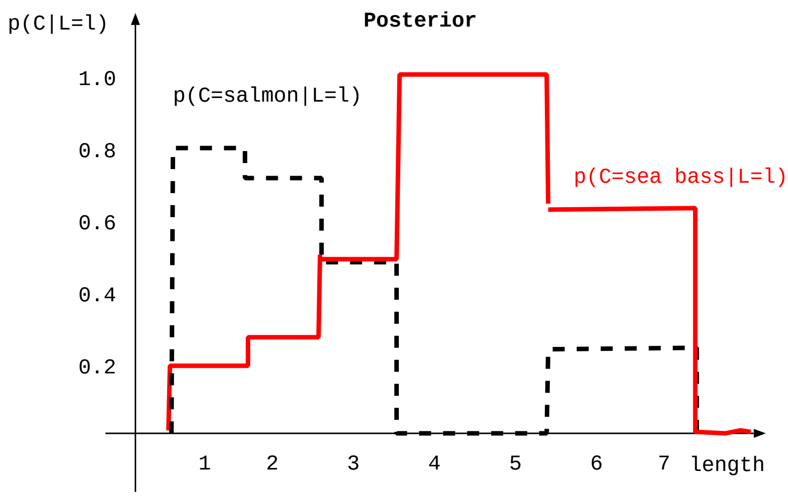
$$P(C = c|L = l) = \frac{P(L = l|C = c)P(C = c)}{P(L = l)} \tag{8}$$

where

- $P(C = c)$ is the prior
- $P(L = l|C = c)$ is the likelihood
- $P(L = l)$ is a normalization constant
- $P(C = c|L = l)$ is the posterior

Note that C is the class (parameter) to be estimated, why L is the data (given).


number of instances
salmon
sea bass
4
3
2
1
1 2 3 4 5 6 7 length
p(L=l|C)
Likelihood
p(L=l|C=salmon)
p(L=l|C=sea bass)
0.4
0.3
0.2
0.1
1 2 3 4 5 6 7 length
p(L=l)
0.25
0.20
0.15
0.10
0.05
1 2 3 4 5 6 7 length


Posterior
p(C|L=l)
1.0
0.8
0.6
0.4
0.2
p(C=salmon|L=l)
p(C=sea bass|L=l)
1
2
3
4
5
6
7
length

How to decide what to do when length=3?

ground truth / classify as	salmon	sea bass
salmon	0	5
sea bass	10	0

Table 1: Loss function

Let R_{10} be the region at which we classify the fish that has length 10. Let $L_{t,c}$ be the loss when classifying a fish of type t as type c. Let H be the length of the fish. Let T be the true class of the fish.

$$T = \begin{cases} s, & \text{salmon} \\ b, & \text{sea bass} \end{cases} \tag{9}$$

If we classify the fish as salmon,

$$\begin{aligned} E[L|H=10] &= E[L|H=10, T=s]p(T=s|H=10) + E[L|H=10, T=b]p(T=b|H=10) = \\ &= L_{s,s}/2 + L_{b,s}/2 = 5 \end{aligned} \tag{10}$$

If we classify the fish as sea bass,

$$\begin{aligned} E[L|H=10] &= E[L|H=10, T=s]p(T=s|H=10) + E[L|H=10, T=b]p(T=b|H=10) = \\ &= L_{s,b}/2 + L_{b,b}/2 = 2.5 \end{aligned} \tag{11}$$

Therefore, the loss is smaller in the second case. We should classify the fish as sea bass.

In general, if we classify the fish in category j,

$$E[L|H=h] = \sum_{k \in \{s,b\}} L_{k,j}p(T=k|H=h) \tag{12}$$

So, given the length of the fish, we have to classify it so as to minimize the above expression.

3 Maximum Entropy

Distributions that maximize entropy

1. among all distributions that have mean μ and standard deviation σ: **normal**

2. among all distributions that have a given support: **uniform**

3. among all distributions that are positive and have mean μ: **exponential**

http://en.wikipedia.org/wiki/Maximum_entropy_probability_distribution

4 Entropy as Minimum Channel Rate