IS 4

Hans Harder

January 21, 2019

1 Derivation of Regression Coefficients

Input: $(x_1, y_1), \ldots, (x_m, y_m)$.

$$\begin{aligned}
f_i(x) &= x^i \\
f(x) &= w_0 + w_1 x + w_2 x^2 + \cdots + w_n x^n \\
&= w_0 f_0(x) + \cdots + w_n f_n(x) \\
&= \sum_{i=1}^{n} w_i f_i(x) \\
\frac{\partial}{\partial w_i} f(x) &= f_i(x) \\
e(y, x) &= y - f(x) \\
&= y - \sum_{i=1}^{n} w_i f_i(x) \\
f^*(x) &= \operatorname*{arg\,min}_{f(x)} \sum_{j=1}^{m} e(y_j, x_j)^2
\end{aligned}$$

a.) Derive the equation that has to be solved for the calculation of w_i.

$$\begin{aligned}
0 &= \frac{\partial}{\partial w_i} \sum_{j=1}^{m} e(x_j)^2 \\
&= \sum_{j=1}^{m} \frac{\partial}{\partial w_i} e(x_j)^2 \\
&= \sum_{j=1}^{m} \frac{\partial}{\partial w_i} (y_j - f(x_j))^2 \\
&= \sum_{j=1}^{m} \frac{\partial}{\partial w_i} y_j^2 - 2y_j f(x_j) + f(x_j)^2 \\
&= \sum_{j=1}^{m} -2y_j f_i(x_j) + \frac{\partial}{\partial w_i} f(x_j)^2 \\
&= \sum_{j=1}^{m} -2y_j f_i(x_j) + 2\left(\frac{\partial f(x_j)}{\partial w_i}\right) f(x_j) \\
0 &= \sum_{j=1}^{m} f_i(x_j) f(x_j) - f_i(x_j) y_j \\
\sum_{j=1}^{m} x_j^i y_j &= \sum_{j=1}^{m} \sum_{k=1}^{n} w_k x_j^{k+i} \\
\sum_{j=1}^{m} x_j^i y_j &= \sum_{j=1}^{m} \left(\sum_{k=1;k\neq i}^{n} w_k x_j^{k+i} \right) + w_i x_j^{2i} \\
\sum_{j=1}^{m} x_j^i y_j &= \sum_{j=1}^{m} \sum_{k=1;k\neq i}^{n} w_k x_j^{k+i} + \sum_{j=1}^{m} w_i x_j^{2i} \\
w_i &= \frac{\sum_{j=1}^{m} x_j^i y_j - \sum_{j=1}^{m} \sum_{k=1;k\neq i}^{n} w_k x_j^{k+i}}{\sum_{j=1}^{m} x_j^{2i}}
\end{aligned}$$

Results:

$$w_i = \frac{\sum_{j=1}^{m} x_j^i y_j - \sum_{j=1}^{m} \sum_{k=1;k\neq i}^{n} w_k x_j^{k+i}}{\sum_{j=1}^{m} x_j^{2i}} \tag{1}$$

b.) Solve the equation above for the case of a straight line ($f_0 = 1, f_1 = x, \forall i > 1 : f_i = 0$).

$$
\begin{aligned}
w_0 &= \frac{\sum_{j=1}^m y_j - w_1 \sum_{j=1}^m x_j}{m} \\
w_1 &= \frac{\sum_{j=1}^m x_j y_j - w_0 \sum_{j=1}^m x_j}{\sum_{j=1}^m x_j^2} \\
&= \frac{\sum_{j=1}^m x_j y_j - \left(\frac{\sum_{j=1}^m y_j - w_1 \sum_{j=1}^m x_j}{m}\right) \sum_{j=1}^m x_j}{\sum_{j=1}^m x_j^2} \\
&= \frac{\sum_{j=1}^m x_j y_j - \left(\sum_{j=1}^m y_j - w_1 \sum_{j=1}^m x_j\right) \sum_{j=1}^m x_j}{m \sum_{j=1}^m x_j^2} \\
&= \frac{\sum_{j=1}^m x_j y_j - \sum_{j=1}^m x_j \sum_{j=1}^m y_j - w_1 \sum_{j=1}^m x_j \sum_{j=1}^m x_j}{m \sum_{j=1}^m x_j^2} \\
&= \frac{\sum_{j=1}^m x_j y_j - \sum_{j=1}^m x_j \sum_{j=1}^m y_j}{m \sum_{j=1}^m x_j^2} - w_1 \frac{\sum_{j=1}^m x_j \sum_{j=1}^m x_j}{m \sum_{j=1}^m x_j^2} \\
&= \frac{\sum_{j=1}^m x_j y_j - \sum_{j=1}^m x_j \sum_{j=1}^m y_j}{m \sum_{j=1}^m x_j^2 \left(1 + \frac{\sum_{j=1}^m x_j \sum_{j=1}^m x_j}{m \sum_{j=1}^m x_j^2}\right)} \\
&= \frac{\sum_{j=1}^m x_j y_j - \sum_{j=1}^m x_j \sum_{j=1}^m y_j}{m \sum_{j=1}^m x_j^2 + \sum_{j=1}^m x_j \sum_{j=1}^m x_j} \\
&= \frac{\sum_{j=1}^m x_j y_j - \sum_{j=1}^m x_j \sum_{j=1}^m y_j}{m \sum_{j=1}^m x_j^2 + \left(\sum_{j=1}^m x_j\right)^2}
\end{aligned}
$$

Results:

$$
w_1 = \frac{\sum_{j=1}^m x_j y_j - \sum_{j=1}^m x_j \sum_{j=1}^m y_j}{m \sum_{j=1}^m x_j^2 + \left(\sum_{j=1}^m x_j\right)^2} \tag{2}
$$

$$
w_0 = \frac{\sum_{j=1}^m y_j - w_1 \sum_{j=1}^m x_j}{m} \tag{3}
$$

2 Linear Regression

s	c
0	0
30	3.5
50	5
80	6.8
100	7.4
130	8
180	12

Compute the hypothesis by means of linear regression.

$$w_1 = 0.06$$
$$w_0 = 1.21$$
$$f(x) = 1.21 + 0.06x$$

3 Gradient Descent

a.) Plot the gradient of the following functions.

$$f(x,y) = \frac{1}{x^2+y^2}$$
$$g(x,y) = x^2 y$$
$$h(x,y) = x + y$$

$$\begin{aligned}
\nabla f(x,y) &= \begin{bmatrix} \frac{\partial}{\partial x} f(x,y) & \frac{\partial}{\partial} y f(x,y) \end{bmatrix} \\
&= \begin{bmatrix} \frac{-2x}{(x^2+y^2)^2} & \frac{-2y}{(x^2+y^2)^2} \end{bmatrix} \\
\nabla g(x,y) &= \begin{bmatrix} 2xy & x \end{bmatrix} \\
\nabla h(x,y) &= \begin{bmatrix} 1 & 1 \end{bmatrix}
\end{aligned}$$

b.) Explain the application of the gradient descent method with respect to one of the functions given in a). The gradient is a vector which points into the direction of greatest increase of the function in respect to its parameter values, in the parameter space (x,y). $h(x,y) = x+y$, for example, increases if x or y increase. The gradient indicates this by having two positive values for x and also y.

c.) (i) Which problems may arise in applying the gradient descent method and how could you tackle them? This method stops as soon as a local minimum is reached, even though a much better minimum could exist. A way to solve this problem would be to do multiple runs with different start values for each parameter.

c.) (ii) Did you run into problems in a)? If yes: which ones, and how did you solve them? Did not really plot the functions (or gradients), but $f(x, y)$ is not defined for some values of x, y, like $x = 0, y = 0$.