Assignment #1

Elements of Machine Learning

Saarland University – Winter Semester 2024/25

Rabin Adhikari
7072310
raad00002@stud.uni-saarland.de

Dhimitrios Duka
7059153
dhdu00001@stud.uni-saarland.de

Problem 3 (Linear Regression)

Derive Residual Sum of Squares (RSS) is the sum of squared residuals for all data points. Make sure to customize it to our model.

First, we will start by defining the Residual Sum of Squares (RSS) formula. Mathematically, RSS is defined as:

$$RSS = \sum_{i=1}^{N} e_i^2 = \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \tag{1}$$

Furthermore, Equation 1 can be expressed in matrix notation as follows:

$$RSS = (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})^{\mathbf{T}} (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \tag{2}$$

For our model, RSS can be expanded as:

$$RSS = \sum_{i=1}^{N} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} \right)^2 \tag{3}$$

Derive and compute the estimates $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2$ that minimize the residual sum of squares by taking the partial derivatives of the RSS with respect to each coefficient.

We need to derive and compute the estimates $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\beta}_2$ that minimize the RSS. Because RSS is a convex problem, we can just take the partial derivatives of the RSS w.r.t each estimate and set it to 0. First, we will start with $\hat{\beta}_0$ since that is also the simplest one.

$$\begin{aligned} \frac{\partial RSS}{\partial \hat{\beta}_0} &= \sum_{i=1}^{N} 2 \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} \right) \cdot (-1) \\ &= -2 \sum_{i=1}^{N} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} \right) \end{aligned} \tag{4}$$

Equating it with zero, we get:

$$
\begin{aligned}
-2\sum_{i=1}^{N}\left(y_i-\hat{\beta}_0-\hat{\beta}_1 x_{i1}-\hat{\beta}_2 x_{i2}\right)&=0\\
\sum_{i=1}^{N} y_i-\sum_{i=1}^{N}\hat{\beta}_0-\hat{\beta}_1\sum_{i=1}^{N} x_{i1}-\hat{\beta}_2\sum_{i=1}^{N} x_{i2}&=0\\
\sum_{i=1}^{N} y_i-N\hat{\beta}_0-\hat{\beta}_1\sum_{i=1}^{N} x_{i1}-\hat{\beta}_2\sum_{i=1}^{N} x_{i2}&=0\\
\frac{1}{N}\sum_{i=1}^{N} y_i-\frac{\hat{\beta}_1}{N}\sum_{i=1}^{N} x_{i1}-\frac{\hat{\beta}_2}{N}\sum_{i=1}^{N} x_{i2}&=\hat{\beta}_0
\end{aligned}
\tag{5}
$$

Next, following a similar procedure, we can derive the estimates for $\hat{\beta}_1$ and $\hat{\beta}_2$. First, we will derive the estimate for $\hat{\beta}_1$.

$$
\begin{aligned}
\frac{\partial RSS}{\partial\hat{\beta}_1}&=\sum_{i=1}^{N} 2\left(y_i-\hat{\beta}_0-\hat{\beta}_1 x_{i1}-\hat{\beta}_2 x_{i2}\right)\cdot(-x_{i1})\\
&=-2\sum_{i=1}^{N} x_{i1}\left(y_i-\hat{\beta}_0-\hat{\beta}_1 x_{i1}-\hat{\beta}_2 x_{i2}\right)
\end{aligned}
\tag{6}
$$

Equating it with zero, we get:

$$
\begin{aligned}
-2\sum_{i=1}^{N} x_{i1}\left(y_i-\hat{\beta}_0-\hat{\beta}_1 x_{i1}-\hat{\beta}_2 x_{i2}\right)&=0\\
\sum_{i=1}^{N} x_{i1}y_i-\sum_{i=1}^{N}\hat{\beta}_0 x_{i1}-\sum_{i=1}^{N}\hat{\beta}_1 x_{i1}^2-\sum_{i=1}^{N}\hat{\beta}_2 x_{i1}x_{i2}&=0
\end{aligned}
\tag{7}
$$

So, we get:

$$
\begin{aligned}
\hat{\beta}_0\cdot\sum_{i=1}^{N} x_{i1}+\hat{\beta}_1\sum_{i=1}^{N} x_{i1}^2+\hat{\beta}_2\sum_{i=1}^{N} x_{i1}x_{i2}&=\sum_{i=1}^{N} x_{i1}y_i\\
\frac{\sum_{i=1}^{N} y_i\, x_{i1}-\hat{\beta}_0\sum_{i=1}^{N}\ x_{i1}-\hat{\beta}_2\sum_{i=1}^{N}\ x_{i1}x_{i2}}{\sum_{i=1}^{N} x_{i1}^2}&=\hat{\beta}_1
\end{aligned}
\tag{8}
$$

Similarly, due to symmetry, we can obtain the optimal value for $\hat{\beta}_2$ also by replacing x_{i1} with x_{i2} and vice versa, $\hat{\beta}_1$ with $\hat{\beta}_2$ and vice versa.

$$
\begin{aligned}
\hat{\beta}_0\cdot\sum_{i=1}^{N} x_{i2}+\hat{\beta}_1\sum_{i=1}^{N} x_{i1}x_{i2}+\hat{\beta}_2\sum_{i=1}^{N} x_{i2}^2&=\sum_{i=1}^{N} x_{i2}y_i\\
\frac{\sum_{i=1}^{N} y_i\, x_{i2}-\hat{\beta}_0\sum_{i=1}^{N}\ x_{i2}-\hat{\beta}_1\sum_{i=1}^{N} x_{i1}x_{i2}}{\sum_{i=1}^{N} x_{i2}^2}&=\hat{\beta}_2
\end{aligned}
\tag{9}
$$

Now, solving equations 5, 8, and 9, we get the optimal values of $\hat{\beta}_0$, $\hat{\beta}_1$, and $\hat{\beta}_2$, for least squares method.

Table 1: The calculation of parameters for the normal equations.

	x_{i1}	x_{i2}	y_i	x_{i1}^2	x_{i2}^2	$x_{i1} \cdot x_{i2}$	$x_{i1} \cdot y_i$	$x_{i2} \cdot y_i$
	1	2	5	1	4	2	5	10
	2	1	6	4	1	2	12	6
	3	3	9	9	9	9	27	27
	4	2	10	16	4	8	40	20
	5	3	13	25	9	15	65	39
Sum	15	11	43	55	27	36	149	102

So, using Table 1 and the normal equations we get the following equations.

$$
\begin{aligned}
43 &= 5\hat{\beta}_0 + 15\hat{\beta}_1 + 11\hat{\beta}_2 \\
149 &= 15\hat{\beta}_0 + 55\hat{\beta}_1 + 36\hat{\beta}_2 \\
102 &= 11\hat{\beta}_0 + 36\hat{\beta}_1 + 27\hat{\beta}_2
\end{aligned}
$$

Solving above equations, we get, $\hat{\beta}_0 = 1.642$, $\hat{\beta}_1 = 1.779$, and $\hat{\beta}_2 = 0.737$.

Compute the R-square value for our model.

We know that R^2 is defined as $R^2 = 1 - \frac{RSS}{TSS}$, where $TSS = \sum_{i=0}^{N} (y_i - \overline{y})^2$ and $RSS = \sum_{i=0}^{N} (y_i - \hat{y}_i)^2$. So, to calculate RSS and TSS let us calculate the necessary values in Table 2.

Table 2: Calculations for R^2 with $\overline{y} = 8.6$

x_{i1}	x_{i2}	y_i	$\hat{y}_i$	$(y_i - \overline{y})^2$	$(y_i - \hat{y})^2$
1	2	5	4.895	12.96	0.011
2	1	6	5.937	6.76	0.0040
3	3	9	9.189	0.16	0.036
4	2	10	10.232	1.96	0.054
5	3	13	12.747	19.36	0.064

From the above table, we get $RSS = 0.168$ and $TSS = 41.2$. Hence, R^2 can be calculated as $R^2 = 1 - \frac{0.168}{41.2} = 0.9959 \approx 1$.