Solution 1:

Which of the following statement(s) is/are correct?

(a) Interpretation methods are mainly needed to better explain real world phenomena. ⇒ **Wrong**

(b) Model-agnostic methods need access to gradients to explain a model. ⇒ **Wrong**

(c) In IML we distinguish between global IML methods, which explain the behavior of the model over the entire feature space, and local IML methods, which only explain the prediction of individual observations. ⇒ **Correct**

(d) We can also draw conclusions about feature importance from feature effect methods. ⇒ **Correct**, but it should be noted that this does not hold vice versa.

(e) Technically, correlation is a measure of *linear* statistical dependence. ⇒ **Correct**

(f) Features that have an equal feature effect are correlated. ⇒ **Wrong**

Solution 2:

a) Calculation of Pearson correlation coefficient of x_1 and x_2

$$\rho(x_1, x_2) = \frac{\sum_{i=1}^{n}(x_1^{(i)} - \overline{x}_1)(x_2^{(i)} - \overline{x}_2)}{\sqrt{\sum_{i=1}^{n}(x_1^{(i)} - \overline{x}_1)^2}\sqrt{\sum_{i=1}^{n}(x_2^{(i)} - \overline{x}_2)^2}}$$

given the dataset

	1	2	3	4	5	6	7	8	9	10	11	$\sum_{i=1}^{n}$
y	-7.90	-6.08	-3.74	-1.18	-1.23	-0.55	0.05	0.88	4.74	2.93	2.55	-9.53
x1	-1.00	-0.80	-0.60	-0.40	-0.20	0.00	0.20	0.40	0.60	0.80	1.00	0
x2	0.95	0.65	0.40	0.07	0.06	0.02	0.02	0.14	0.34	0.60	0.98	4.23

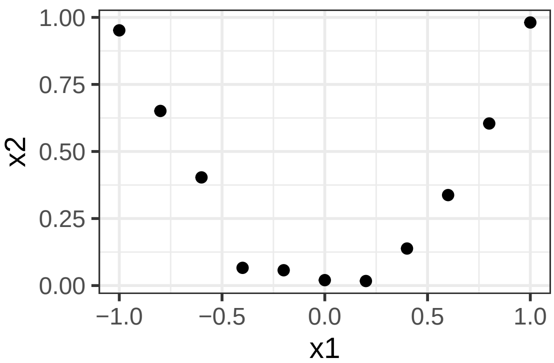
The individual differences to the means are

	1	2	3	4	5	6	7	8	9	10	11
$x_1^{(i)} - \overline{x}_1$	-1.00	-0.80	-0.60	-0.40	-0.20	0.00	0.20	0.40	0.60	0.80	1.00
$x_2^{(i)} - \overline{x}_2$	0.57	0.27	0.02	-0.31	-0.32	-0.36	-0.36	-0.24	-0.04	0.22	0.6

$$\begin{aligned}\rho(x_1, x_2) &= \frac{\sum_{i=1}^{n}(x_1^{(i)} - \overline{x}_1)(x_2^{(i)} - \overline{x}_2)}{\sqrt{\sum_{i=1}^{n}(x_1^{(i)} - \overline{x}_1)^2}\sqrt{\sum_{i=1}^{n}(x_2^{(i)} - \overline{x}_2)^2}} \\ &= \frac{-0.57 + -0.22 + -0.01 + 0.12 + -0.06 + 0 + -0.07 + -0.1 + -0.02 + 0.18 + 0.6}{2.41} = \frac{-0.03}{2.41} = -0.01\end{aligned}$$

The Pearson correlation coefficient is close to 0 ⇒ there is no **linear** relationship between x_1 and x_2.

b) The scatter plot reveals that there is a strong non-linear/quadratic relationship between x_1 and x_2. The Pearson correlation coefficients is not suitable for detecting non-linear relationships.



Solution 3:

First, recall that the formula for the coefficient of determination R^2 is:

$$R^2 = 1 - \frac{SSE_{LM}}{SSE_c}$$

where $SSE_{LM} = \sum_{i=1}^{n}(y^{(i)} - \hat{f}_{LM}(x^{(i)}))^2$ is the sum of squares due to regression and $SSE_c = \sum_{i=1}^{n}(y^{(i)} - \bar{y})^2$ is the total sum of squares. The formula for the Pearson correlation coefficient ρ is:

$$\rho(x,y) = \frac{cov(x,y)}{\sigma_x \sigma_y} = \frac{\sum_{i=1}^{n}(x^{(i)} - \bar{x}) \cdot (y^{(i)} - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x^{(i)} - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y^{(i)} - \bar{y})^2}}.$$

We have:

$$\begin{aligned}
R^2 &= 1 - \frac{SSE_{LM}}{SSE_c} \\
&= 1 - \frac{\sum_{i=1}^{n}(y^{(i)} - \hat{f}_{LM}(x^{(i)}))^2}{\sum_{i=1}^{n}(y^{(i)} - \bar{y})^2} \\
&= 1 - \frac{\sum_{i=1}^{n}(y^{(i)} - \hat{y}^{(i)})^2}{\sum_{i=1}^{n}(y^{(i)} - \bar{y})^2}
\end{aligned}$$

Similarly, we can write:

$$\begin{aligned}
\rho^2 &= \left(\frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y}\right)^2 \\
&= \left(\frac{\sum_{i=1}^{n}(x^{(i)} - \bar{x})(y^{(i)} - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x^{(i)} - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y^{(i)} - \bar{y})^2}}\right)^2 \\
&= \frac{\left(\sum_{i=1}^{n}(x^{(i)} - \bar{x})(y^{(i)} - \bar{y})\right)^2}{\left(\sum_{i=1}^{n}(x^{(i)} - \bar{x})^2\right)\left(\sum_{i=1}^{n}(y^{(i)} - \bar{y})^2\right)} \\
&= \frac{\left(\sum_{i=1}^{n}(x^{(i)} - \bar{x})(y^{(i)} - \bar{y})\right)^2}{\left(\sum_{i=1}^{n}(x^{(i)} - \bar{x})^2\right)\left(\sum_{i=1}^{n}(y^{(i)} - \bar{y})^2\right)} \frac{\sum_{i=1}^{n}(x^{(i)} - \bar{x})^2}{\sum_{i=1}^{n}(x^{(i)} - \bar{x})^2} \\
&= \left(\frac{\sum_{i=1}^{n}(x^{(i)} - \bar{x})(y^{(i)} - \bar{y})}{\sum_{i=1}^{n}(x^{(i)} - \bar{x})^2}\right)^2 \frac{\sum_{i=1}^{n}(x^{(i)} - \bar{x})^2}{\sum_{i=1}^{n}(y^{(i)} - \bar{y})^2} \\
&= \hat{\beta}_1^2 \frac{\sum_{i=1}^{n}(x^{(i)} - \bar{x})^2}{\sum_{i=1}^{n}(y^{(i)} - \bar{y})^2}
\end{aligned}$$

Now, note that

$$\sum_{i=1}^{n}(y^{(i)}-\bar{y})^2=\sum_{i=1}^{n}(y^{(i)}-\hat{y}^{(i)})^2+\sum_{i=1}^{n}(\hat{y}^{(i)}-\bar{y})^2.$$

Proof:

$$\begin{aligned}\sum_{i=1}^{n}(y^{(i)}-\bar{y})^2 &= \sum_{i=1}^{n}(y^{(i)}-\hat{y}^{(i)}+\hat{y}^{(i)}-\bar{y})^2\\ &= \sum_{i=1}^{n}(y^{(i)}-\hat{y}^{(i)})^2+(\hat{y}^{(i)}-\bar{y})^2+2(y^{(i)}-\hat{y}^{(i)})(\hat{y}^{(i)}-\bar{y})\\ &= \sum_{i=1}^{n}(y^{(i)}-\hat{y}^{(i)})^2+\sum_{i=1}^{n}(\hat{y}^{(i)}-\bar{y})^2+2\sum_{i=1}^{n}(y^{(i)}-\hat{y}^{(i)})(\hat{y}^{(i)}-\bar{y})\end{aligned}$$

It remains to show that

$$\begin{aligned}2\sum_{i=1}^{n}(y^{(i)}-\hat{y}^{(i)})(\hat{y}^{(i)}-\bar{y}) &= 0\\ \sum_{i=1}^{n}(y^{(i)}-\hat{y}^{(i)})\hat{y}^{(i)}-\sum_{i=1}^{n}(y^{(i)}-\hat{y}^{(i)})\bar{y} &= 0\\ \bar{y}\sum_{i=1}^{n}y^{(i)}-\hat{y}^{(i)} &= 0\\ \sum_{i=1}^{n}y^{(i)}-\hat{y}^{(i)} &= 0\end{aligned}$$

where we have used the fact that $\sum_{i=1}^{n}(y^{(i)}-\hat{y}^{(i)})\hat{y}^{(i)}=0$ as the residuals $(y^{(i)}-\hat{y}^{(i)})$ and $\hat{y}^{(i)}$ are not correlated. Substituting these results into the expression for R^2, we obtain:

$$\begin{aligned}R^2 &= 1-\frac{\sum_{i=1}^{n}(y^{(i)}-\hat{y}^{(i)})^2}{\sum_{i=1}^{n}(y^{(i)}-\bar{y})^2}\\ &= \frac{\sum_{i=1}^{n}(y^{(i)}-\bar{y})^2-\sum_{i=1}^{n}(y^{(i)}-\hat{y}^{(i)})^2}{\sum_{i=1}^{n}(y^{(i)}-\bar{y})^2}\\ &= \frac{\sum_{i=1}^{n}(y^{(i)}-\hat{y}^{(i)})^2+\sum_{i=1}^{n}(\hat{y}^{(i)}-\bar{y})^2-\sum_{i=1}^{n}(y^{(i)}-\hat{y}^{(i)})^2}{\sum_{i=1}^{n}(y^{(i)}-\bar{y})^2}\\ &= \frac{\sum_{i=1}^{n}(\hat{y}^{(i)}-\bar{y})^2}{\sum_{i=1}^{n}(y^{(i)}-\bar{y})^2}\\ &= \frac{\sum_{i=1}^{n}(\hat{\beta}_0+\hat{\beta}_1x^{(i)}-(\hat{\beta}_0+\hat{\beta}_1\bar{x}))^2}{\sum_{i=1}^{n}(y^{(i)}-\bar{y})^2}\\ &= \frac{\hat{\beta}_1^2\sum_{i=1}^{n}(x^{(i)}-\bar{x})^2}{\sum_{i=1}^{n}(y^{(i)}-\bar{y})^2}\\ &= \rho^2\end{aligned}$$

Hence, we have shown that $R^2=\rho^2$, which completes the proof. Note that this result is valid only for simple linear regression, where there is only one independent variable. For multiple regression, the coefficient of determination is defined differently and does not necessarily equal the square of the Pearson correlation coefficient.

Solution 4:

Problem: The function $f(\mathbf{x})=2x_1+3x_2-x_1|x_2|$ is not differentiable for $x_2=0$. Hence, different cases need to be considered:

Case 1: $x_2 > 0$
Case 2: $x_2 < 0$
Case 3: $x_2 = 0$

Case 1: $x_2 > 0$

$$\left(\frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_2}\right)^2 = \left(\frac{\partial^2}{\partial x_1 \partial x_2}\,(2x_1 + 3x_2 - x_1 x_2)\right)^2 = \left(\frac{\partial}{\partial x_2}\,(2 - x_2)\right)^2 = (-1)^2 = 1 > 0$$

Case 2: $x_2 < 0$

$$\left(\frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_2}\right)^2 = \left(\frac{\partial^2}{\partial x_1 \partial x_2}\,(2x_1 + 3x_2 - x_1(-x_2))\right)^2 = \left(\frac{\partial}{\partial x_2}\,(2 + x_2)\right)^2 = 1^2 = 1 > 0$$

Case 3: $x_2 = 0$
Not considered, as analysis of interactions via definition requires the consideration of intervals. The examination of single points does not make sense.

$\Rightarrow$ x_1 and x_2 interact with each other.