2022-05-18 Wednesday

NAME: Ravi Moelchand - 7463332

Topics covered since the last tutorial:

- Linear Models 2 (Linear Regression) (Lecture 5)
- Shattering and VC Dimension (Lecture 6)

And we will also cover this mathematics topic not previously covered in the tutorials:

- Adding probabilities, conditional probabilities, joint probabilities (Lectures 3, 4)

If you write your answers directly into the notebook, it is preferred that you generate a .pdf file for submission

Linear Regression

A music company wants us to predict how much people enjoy listening to a new form of experimental music. They carry out a small poll in five listeners of varying ages and provide us with the following table:

x	-0.9	-0.5	0.3	0.4	0.7
y	6	4	2	2	1

Here, y is the listeners' grading on a scale between 1 (lowest) and 6 (highest), and x represents their ages after normalization: x = (age - 50)/40.

1. What is the age of the oldest participant in the poll?

answer

$x = (age - 50)/40$

For the oldest person in the group of listeners we will have to look at the highest x, which is 0.7.\ So,
$0.7 = (age - 50)/40 \rightarrow 28 = age - 50 \rightarrow age = 28 + 50 = 78$

2. We want to predict grade (y) from age (x) using linear regression: $y = w_0 + w_1 x$. Write down the formula for the in-sample error E_{in} in terms of w_0 and w_1.

answer

$E_{in}(h) = E_{in}(w_0, w_1) = \frac{1}{N}\sum_{i+1}^{N}(w^T x_i - y_i)^2$

3. We have to find the optimal w_0 and w_1 yielding the minimal E_{in}. Someone has told us that, for a dataset with $mean(x) = 0$, $mean(y) = w_0$. What is w_0?

answer

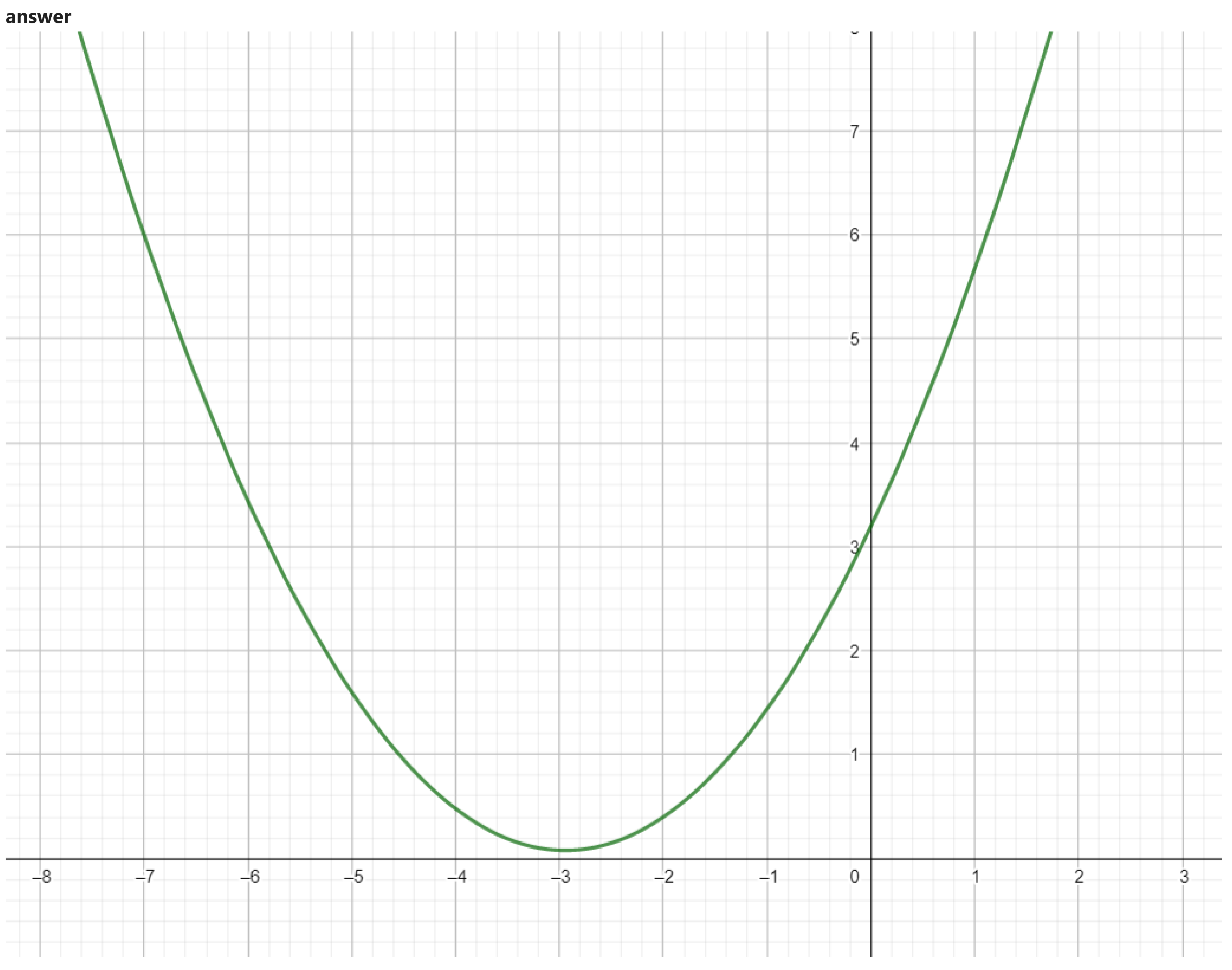
$mean(y) = \frac{6+4+2+2+1}{5} = 3 = w_0$

4. Using this value for w_0, we find that $E_{in}(w_1) = 3.2 + 2.12w_1 + 0.36w_1^2$. (Make sure you can do this yourself too). We can find the optimal value of w_1 and the corresponding E_{in} by taking the derivative of this equation. What is the optimal value of w_1 and the corresponding E_{in} (use 3 decimals precision).

answer
$E_{in}(w_1) = 3.2 + 2.12w_1 + 0.36w_1^2 \rightarrow derivative : E'_{in}(w_1) = 2.12 + 0.72w_1$
$E'_{in}(w_1) = 0.72w_1 + 2.12 = 0 \rightarrow 0.72w_1 = -2.12 \rightarrow w_1 = \frac{-53}{18}$
$E_{in}(\frac{-53}{18}) = 3.2 + 2.12(\frac{-53}{18}) + 0.36(\frac{-53}{18})^2 \approx 0.079$

5. Make a plot of $E_{in}(w_1)$.

answer



6. We don't believe that person, so we will now do it in the official way. Write down matrix X for our problem (don't forget to include the constant).

answer

$$\begin{pmatrix} 1 & -0.9 \\ 1 & -0.5 \\ 1 & 0.3 \\ 1 & 0.4 \\ 1 & 0.7 \end{pmatrix}$$

7. Calculate the least-squares estimator $\mathbf{w}_{\text{lin}}$ ($\mathbf{w}_{\text{lin}} = (\mathrm{X}^T\mathrm{X})^{-1}\mathrm{X}^T\mathbf{y}$).

Hint: In order to do the matrix inversion, note that $\mathrm{A}^{-1}\mathrm{A} = \mathbf{I}$ (the identity matrix) and that the inverse of a diagonal matrix is also diagonal.

answer

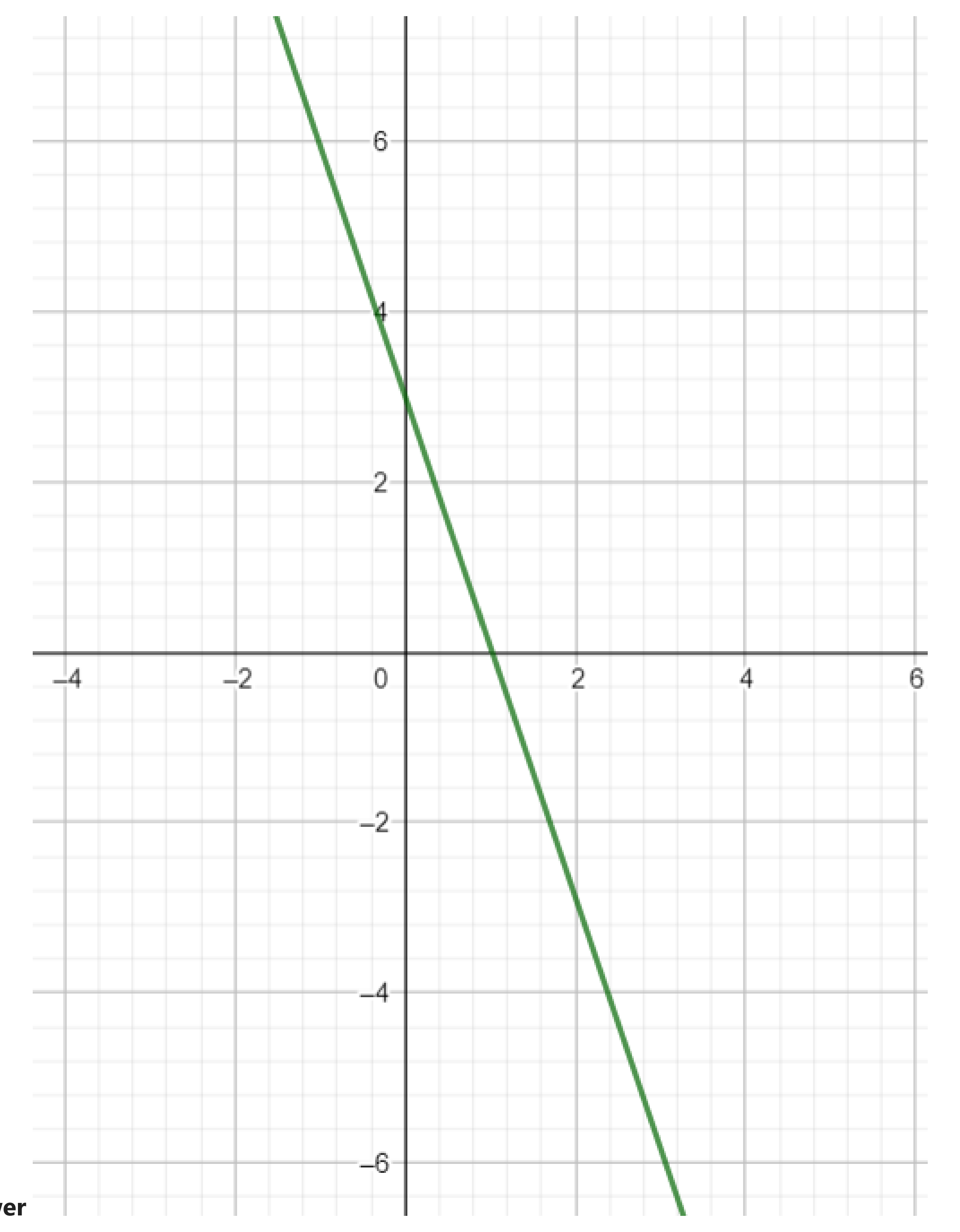
$$\mathbf{w}_{\text{lin}} = (\mathrm{X}^T\mathrm{X})^{-1}\mathrm{X}^T\mathbf{y} = \left(\begin{pmatrix} 1 & -0.9 \\ 1 & -0.5 \\ 1 & 0.3 \\ 1 & 0.4 \\ 1 & 0.7 \end{pmatrix} \cdot \begin{pmatrix} -0.9 & -0.5 & 0.3 & 0.4 & 0.7 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix}\right)^{-1} \cdot \begin{pmatrix} -0.9 & -0.5 & 0.3 & 0.4 & 0.7 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix}\begin{pmatrix} 6 \\ 4 \\ 2 \\ 2 \\ 1 \end{pmatrix}$$
$$= \begin{pmatrix} -1.8 & -1.4 & -0.6 & -0.5 & -0.2 \\ -1.4 & -1 & -0.2 & -0.1 & 0.2 \\ -0.6 & -0.2 & 0.6 & 0.7 & 1 \\ -0.5 & -0.1 & 0.7 & 0.8 & 1.1 \\ -0.2 & 0.2 & 1 & 1.1 & 1.4 \end{pmatrix}^{-1} \begin{pmatrix} -0.9 & -0.5 & 0.3 & 0.4 & 0.7 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix}\begin{pmatrix} 6 \\ 4 \\ 2 \\ 2 \\ 1 \end{pmatrix}$$
$$= \begin{pmatrix} -1.8 & -1.4 & -0.6 & -0.5 & -0.2 \\ -1.4 & -1 & -0.2 & -0.1 & 0.2 \\ -0.6 & -0.2 & 0.6 & 0.7 & 1 \\ -0.5 & -0.1 & 0.7 & 0.8 & 1.1 \\ -0.2 & 0.2 & 1 & 1.1 & 1.4 \end{pmatrix}^{-1} \begin{pmatrix} -5.3 \\ 15 \end{pmatrix} = \begin{pmatrix} 3 \\ -2.944 \end{pmatrix}$$

8. Write down your solution in the form $y = w_0 + w_1 x$. Was our advisor right about w_0? Compare this solution to the one before.

answer

$y = 3 - 2.944x$ \ our advisor was right, because $w_0 = 3$

9. Make a plot of the data and the hypothesis function.



answer

10. Our hypothesis function does not match the training data. Give two reasons why this could be the case.

answer

- *noise factor*
- *rounding*

11. While handing in our prediction model to the music company, we would like to provide them with an estimate of the uncertainty in our predictions. We assume the data were "generated" as $y(x) = f(x) + \epsilon$, with ϵ a noise factor that is randomly drawn from a normal distribution with mean = 0 and variance = σ^2. To calculate this variance, i.e., E_{out}, we would need to know f, while we have only our best hypothesis g at hand. It has been shown that the variance can be best estimated as $N/(N-1) * E_{in}$. Provide an estimate of σ (using the value of E_{in} you found).

answer

12. Does this value make sense? (have a look at the relevant plot in the previous subquestions). Would you consider this noise small or large, as compared to the effect of age on $\mathbf{y}$?

answer

Optional (advanced) Solve the regression problem using the hat matrix. $\hat{\mathbf{y}} = \mathrm{H}\mathbf{y}$; $\mathrm{H} = \mathrm{X}(\mathrm{X}^{\mathrm{T}}\mathrm{X})^{-1}\mathrm{X}^{\mathrm{T}}$. (Calculate H.)

your answer here

Optional (advanced) Examine some of the interesting properties of H in Ex. 3.3 in the book. For (a) use $(\mathrm{A}^{-1})^{\mathrm{T}} = (\mathrm{A}^{\mathrm{T}})^{-1}$. For (b) and (c), start with $K = 2$ and use induction. Verify that (a) and (d) hold for "our" H.

your answer here

Probability

You may want to refresh the foundations of probability theory. Please ask one of the teaching staff if you need extra help with this subject.

13. $\mathbb{P}[\mathcal{A} \textbf{ or } \mathcal{B}]$ is the probability that at least one of $\mathcal{A}$ and $\mathcal{B}$ happens. Show that $\mathbb{P}[\mathcal{A} \textbf{ or } \mathcal{B}] \geq \mathbb{P}(\mathcal{A}) + \mathbb{P}(\mathcal{B}) - 1$.

Hint: Think about the relationship between this question and Hoeffding's inequality.

answer

$P(\mathcal{A})$ and $P(\mathcal{B})$, are both always numbers between 0 and 1.

Suppose that $P(\mathcal{A})$ and $P(\mathcal{B})$ both happen, that means $P(\mathcal{A}) + P(\mathcal{B}) = 1$. That gives us, $\mathbb{P}(\mathcal{A}) + \mathbb{P}(\mathcal{B}) - 1 = 0 \rightarrow 1 > 0$.

Now consider the situation where neither happen, that means $P(\mathcal{A}) + P(\mathcal{B}) = 0$. Which gives us, $\mathbb{P}(\mathcal{A}) + \mathbb{P}(\mathcal{B}) - 1 = -1 \rightarrow 0 > -1$

In every other situation, $0 < P(\mathcal{A}) + P(\mathcal{B}) < 1$. So the outcome of $(\mathcal{A}) + \mathbb{P}(\mathcal{B}) - 1$ will always be a negative number greater than -1. Therefore $P[\mathcal{A} \textbf{ or } \mathcal{B}] \geq (\mathcal{A}) + \mathbb{P}(\mathcal{B}) - 1$

14. There are three boxes, each with two drawers. Box 1 has a gold coin in each drawer, and box 2 has a silver coin in each drawer. Box 3 has a silver coin in one drawer and a gold coin in the other. One box is chosen at random, then a drawer is chosen at random from the box. Find the probability that box 1 is chosen, given that the chosen drawer yields a gold coin.

your answer here