Machine Learning

ML Quiz #3

Mostafa S. Ibrahim
Teaching, Training and Coaching for more than a decade!

Artificial Intelligence & Computer Vision Researcher
PhD from Simon Fraser University - Canada
Bachelor / MSc from Cairo University - Egypt
Ex-(Software Engineer / ICPC World Finalist)

Solutions Submissions

- Only submit BEFORE you listen to the answers during the lecture
- Submission form

Questions A: True or False

1. Polynomial regression can be used to model **nonlinear** relationships between variables by building *polynomial relationship* between the independent and dependent variables
2. Polynomial regression can't capture **interactions** between variables
3. In polynomial regression, the degree of the polynomial determines the **complexity** of the model and subjective to overfitting when the degree of the polynomial is too high
4. I.i.d. data is a common assumption that holds in **practice**
5. When comes to data split into train/val/test, the key objective is to strike a balance between having enough data for training, validating the model's performance, and having a reliable estimate of its generalization ability on unseen data.

Questions B: True or False

1. With k-fold cross-validation, all data points are tried for both training and validation
2. K-fold cross-validation is used for hyperparameter tuning and model selection
3. In practice, we always select the model with the lowest validation error
4. Overfitting can be reduced by increasing the model's complexity
5. Underfitting can be reduced by providing more training data

Questions C

1. Why does k-fold cross-validation provide a **more reliable estimate** of the model's performance compared to a single train-test split?
2. How can cross-validation detect if a model is overfitting the training data?
3. In market analysis, data may be collected from different market segments, regions, or customer segments. Each segment can have unique characteristics, behaviors, or preferences. Assume we want to train our model so that it works well for **unseen segments** of customers. Which specific cross validation technique should be used. Explain

Question D: Fix the code

In this code an ellipse is created

Then, we transform it into line using x^2, y^2 transformation

However, the code still shows an ellipse

Write a fixed version of:
From_ellipse_to_line

- It should be generate solution, not to just these data

```
def from_ellipse_to_line(x, y):
    return x ** 2, y ** 2

cx, cy = -10, 10    # center
rx, ry = 5, 5       # radius
size = 10000

t = np.linspace(0, 2*pi, size)
x = cx+rx*np.cos(t)
y = cy+ry*np.sin(t) + np.random.normal(0, 0.2, size)

x, y = from_ellipse_to_line(x, y)

plt.scatter(x, y)
plt.grid()
plt.show()
```