Solutions Exam Computational Statistics

Q1: 2.16

Q2: Points 1 and 4 are in the training set, so the model is a straight line through those two points. Hence, the validation MSE is 1/3 * (2^2+2^2+4^2) = 1/3 * 24 = 8

Q3: "The error of the cubic regression to be lower."

Q4: "The error of the linear regression to be lower."

Q5: "The error of the cubic regression to be lower."

Q6: "There is not enough information to tell which regression model should have lower error."

Q7: 8

Q8: 6.5

Q9: 2

Q10: False

Q11: False

Q12: False

Q13: False

Q14: True

Q15: False

Q16: model, X, y

Q17: predict_proba

Q18: :,1

Q19: .mean()

Q20: targets

Q21: iloc

Q22: features

Q23: iloc

Q24: rf = RandomForestClassifier()

Q25: rf.fit(X_train, y_train)

Q26: X_validate, y_validate

Q27: 1/3

Q28: 1.0

Q29: Any of these

Q30: 0.4

Q31: f_1 = 1/3 + 1/3 * (1-(0.5-1)^2 - 1/2*(1-(1-1)^2 + 1-(1-1)^2))
= 1/3 + 1/3*(0.75 - 1/2*2)
= 1/3 + 1/3*(-1/4)
= 0.25

Q32: 0.825

Q33: 0.4286

Q34: Expected score not knowing H: 0.75*(1-(0.75-1)^2) + 0.25*(1-(0.75-0)^2) = 0.8125

Q35: Expected score not knowing H: 0.8125
Expected score after learning H: 0.825 * 0.851 + 0.175 * 0.755 = 0.8342
Expected improvement: 0.8342 - 0.8125 = 0.0217