WORKSHEET

U12.3 Generalisation

Learning, and the capstone · University · about 35 min

Name _____ Class _____ Date _____

What this lesson is about

Held-out data, overfitting, and choosing a model by the error it makes on points it never saw.

- The claim a fit makes
- Overfitting, concretely
- The only honest test
- Fitting a polynomial in plain Python
- Model selection

Questions

7 marks in all

1. A degree 5 polynomial through six noisy speed measurements has zero training error. What has it learned? [1 mark]

- ☐ **A** Mainly the noise in those six points, so it cannot be trusted on new commands, least of all beyond the data
- ☐ **B** The true relationship exactly
- ☐ **C** The same as a straight line, since both fit the data
- ☐ **D** Nothing, because a zero error means the fit failed

2. A line is fitted to the three training points and scored by RMS error on both sets. What does this print? [1 mark]

```
train = [(30, 5.8), (50, 9.9), (70, 14.1)]
test = [(40, 8.2), (60, 11.7), (80, 16.3)]
n = len(train)
sx = sum(c for c, v in train)
sv = sum(v for c, v in train)
sxx = sum(c * c for c, v in train)
sxv = sum(c * v for c, v in train)
a = (n * sxv - sx * sv) / (n * sxx - sx * sx)
b = (sv - a * sx) / n

def rms(points):
    return (sum((a * c + b - v) ** 2 for c, v in points) / len(points)) ** 0.5

print(round(rms(train), 3), round(rms(test), 3))
```

3. A student compares five models by their test error, picks the best, and reports that same test error as the result. What is wrong? [1 mark]

- ☐ **A** The test set has been used to choose the model, so the reported error is optimistic; a separate validation set should do the choosing
- ☐ **B** Nothing, as long as the test points were never used to fit
- ☐ **C** Five models is too few to compare
- ☐ **D** Test error should be reported on the training data instead

4. With very little data, what method splits the data into k parts, fits k times leaving each part out in turn, and averages the held-out errors? [1 mark]

5. Which statements about bias and variance are right? [1 mark]

Tick every answer that is true.

- ☐ **A** A model with too few parameters is wrong the same way every time
- ☐ **B** A model with too many parameters is wrong a different way each time it is refitted on fresh data
- ☐ **C** The error on new data has a minimum somewhere in between
- ☐ **D** Adding parameters always lowers the error on new data
- ☐ **E** Training error measures variance

6. Why scale commands, u = (c - 55) / 25, before fitting a high degree polynomial by the normal equations? [1 mark]

- ☐ **A** Raw commands raised to high powers differ by many orders of magnitude, and the solution comes back as noise
- ☐ **B** Scaling improves the model's held-out error
- ☐ **C** The normal equations only accept values between 0 and 1
- ☐ **D** It removes the dead band from the data

7. Two models score nearly the same held-out error, one degree 1 and one degree 3. Which should you choose, and why? [1 mark]

- ☐ **A** Degree 1, because the simpler model has less variance, so its held-out score is more trustworthy
- ☐ **B** Degree 3, because it fits the training data better
- ☐ **C** Degree 3, because it can represent more shapes
- ☐ **D** Either, because held-out error is the only thing that matters

The task: choose the model by held-out error

Sample fourteen commands, six for training and eight for testing, fit a degree 1 and a degree 5 model to the training six, print all four errors, and print `train 5:` and `best degree:`. Three of the test commands, 25, 85 and 95, lie outside the training range, which is where an overfitted model is at its worst.

```
from bugbot import *
connect()

DT = 0.1
TRAIN = [30, 40, 50, 60, 70, 80]
TEST = [25, 35, 45, 55, 65, 75, 85, 95]
```

Plan your program here, then type it in and press Run.

Do it on the robot

www.bugbotlab.com/learn/u12-3-generalisation/

The simulator checks it and tells you when it passes. Nothing to install, no account.

Challenges

1. Print the degree 5 model's prediction at command 120. What has it done, and why is that the expected behaviour?
2. Try degrees 1, 2 and 3 as well and plot training and test error against degree.
3. Swap your train and test sets over and refit. How different is the fitted slope, and what does that difference tell you?