Course / UNIT 4 / Problem Set 4

Previous | Next

Problem 3

🔖 Bookmark this page

Calculator   ✏ Hide Notes

Problem 3

5.0/5.0 points (graded)

We have learned how to obtain a numerical metric for evaluation. Visualizing our data samples along with fitting curves can also help us figure out the goodness of obtained models. In this problem, we will integrate the numerical metrics and visualization for a comprehensive evaluation.

Implement function `evaluate_models_on_training`. This function takes as input your data samples (x and y) and the list of models (which are lists of coefficients obtained from `generate_models`) that you want to apply to your data.

This function should generate a figure for each model. In this figure, you are to plot your data along with your best fit curve, and report on the goodness of the fit with the R^2 value. When you are writing this function try to make your graph match the following format:

- Plot the data points as individual blue dots

- Plot your model as a red solid line

- Include a title and label your axes

- Your title should include the value of your model and the R^2 degree of this model. Your title could be longer than your graph. To fix that you can add "\n", which adds a newline to your string, in your title when you concatenate several pieces of information (e.g., title = string_a + "\n" + string_b).

After you finish writing the function, you have all the components needed to start generating data samples from the raw temperature records and investigate the trend. Run the following code at the bottom ps4.py.

```
# Problem 3
y = []
x = INTERVAL_1
for year in INTERVAL_1:
    y.append(raw_data.get_daily_temp('BOSTON', 1, 10, year))
models = generate_models(x, y, [1])
evaluate_models_on_training(x, y, models)
```

This code just randomly picks a day from a year (i.e., Jan 10th in this case), and sees whether we can find any trend in the temperature changing over the years. We surmise, due to global warming, that the temperature of this specific date should increase over time. This code generates your data samples; each sample represents a year from 1961 to 2005 (i.e., the years in `INTERVAL_1`) and the temperature of Jan 10th for Boston in that year (provided helper class is helpful for this). The code fits your data to a linear line with `generate_models` and plots the regression results with `evaluate_models_on_training` .

What is the R^2 value? (use 3 decimal places)

```
0.005
```
✔

0.005

Submit

Problem 3

Hide Discussion

Topic: Problem Set 4 / Problem 3

Show all posts ⌄   by recent activity ⌄

| ? | **Does it seem like there's a lot missing in this course?** | 8 |
| | It seems like it is barely possible to answer these PSETs without doing a lot of outside googling / youtubing to find better methods. I... | |

| ? | **my data points are all over the place** | 3 |
| | my r2 is correct, so either its ok, or im doing something wrong in the plot. anything I should be paying atten... [⊞ Calculator] ... | |

edX

About

Affiliates

edX for Business

Open edX

Careers

News

Legal

Terms of Service & Honor Code

Privacy Policy

Accessibility Policy

Trademark Policy

Sitemap

Connect

Blog

Contact Us

Help Center

Security

Media Kit

Calculator