

homework_06

May 11, 2020

0.1 Exercise 1

Poisson regression is a Generalized Linear Model, used to model count data. It takes the form

$$\mathbb{E}(\mu|x) = \exp(w_1 x_1 + \dots + w_k x_k + b),$$

where the observed counts y are drawn from a Poisson distribution on the expected counts:

$$y_i \sim \text{Poisson}(\mu_i).$$

1. Download and import Load the smoking dataset from: <https://data.princeton.edu/wws509/datasets/#smoking>. Then perform a train-test split on the data;
2. Fit a Poisson bayesian regression model using the number of deaths as the response variable and the other columns as the explanatory variables, using Pyro Variational Inference;
3. Evaluate the regression fit on test data using MAE and MSE error metrics.

0.2 Exercise 2

The Iris dataset contains petal and sepal length and width for three different types of Iris flowers: Setosa, Versicolour, and Virginica.

1. Import the Iris dataset from `sklearn`:

```
from sklearn import datasets
iris = datasets.load_iris()
```

and perform a train-test split on the data.

2. Fit a multinomial bayesian logistic regression model on the four predictors petal length/width and sepal length/width, using Pyro VI.
3. Evaluate your bayesian classifier on test data: compute the overall test accuracy and class-wise accuracy for the three different flower categories.

[]: