ICS 435/635

Writing

Writing a ML Methods Section

ML Methods Sections

1. Introduce the problem
2. Introduce the data
3. Introduce the model

Convince the reader this is a good use of ML.

4. Specify data splitting
5. Specify data pre-processing
6. Specify hyperparameters
7. Specify hyperparameter optimization

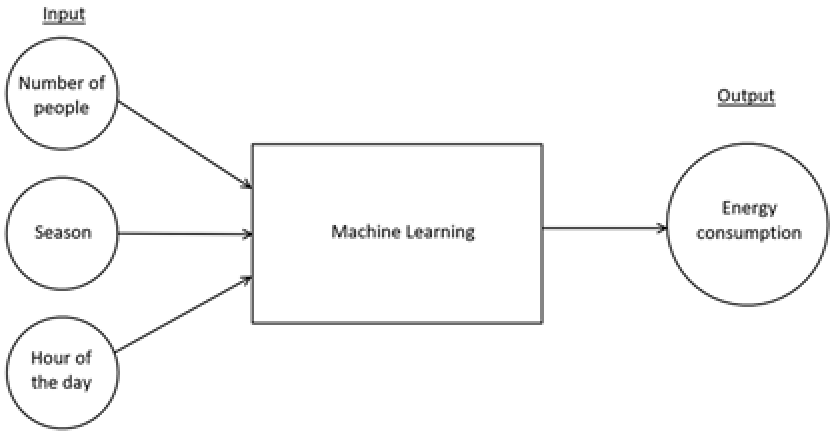
Enable reader to replicate your results.

8. Demonstrate performance on clean test set
9. Discuss generalization

Convince the reader your model performance will generalize.

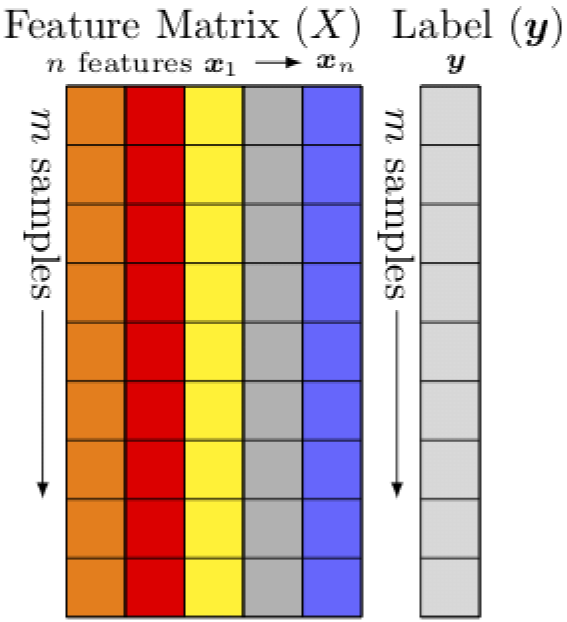
Introduce the Problem

- What problem are you trying to solve?
- What will model be used for?
- What are the inputs and outputs?
- What function are you approximating?
- Why is ML appropriate for this task?
- Why do other methods fail?
- Where is the training data coming from?



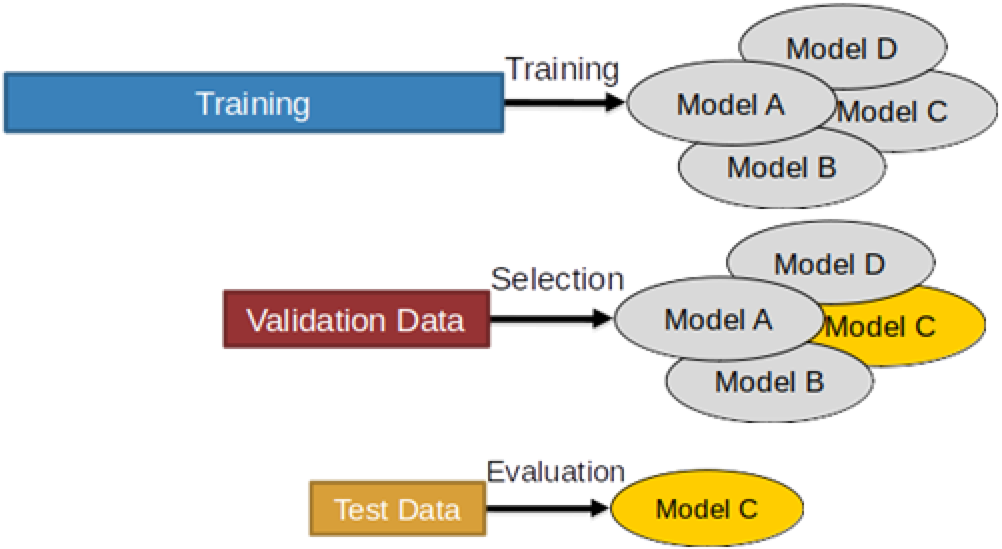
Introduce the Data

- How many samples are there?
- How much labelled/unlabelled data?
- Do you trust the labels?
- Will the data size increase in the future?



Specify Data Splitting

- Test set is necessary for evaluation.
- Validation set is necessary for model selection / hyperparameter optimization.
- Specify upfront so reader knows you are careful.



Specify Data Pre-processing

- Feature units
- Feature properties
  - Categorical, ordinal, one-hot
- Transformations
  - Min-max scaling, standardization, log(1+x)
- Missing data methods
  - Removal, imputation,
- Augmentation

Specify Hyperparameters

- Model selection
- Regularization (always)
- Training algorithm

Specify Hyperparam Optimization

Algorithms:

- Manual
- Random Search
- Grid Search
- Bayesian Optimization

Tools:

- Optuna
- Sherpa

Demonstrate Performance on Clean Test Set

- Every number should be accompanied by the **data split** and **metric**.
- It is common to compare different models on final test data.
  - This means most research papers over-estimate performance!

Example: On the clean test set, our optimized DT got 0.87 test AUROC, while the optimized K-NN classifier got only 0.85 test AUROC. The DT's performance was much lower than on the validation set (0.92 validation AUROC, 95% validation accuracy), which suggests either overfitting to the validation set or differences in the validation and test sets.

Evaluation: Common Mistakes

K-NN got a score of 0.96. ---> what is “score”?

K-NN got a AUROC of 0.96, and the test accuracy was 90%. ---> AUROC on what dataset?

K-NN got a test set AUROC of 0.96822. ---> too many significant digits.

K-NN was used to make predictions. ---> What value of K?

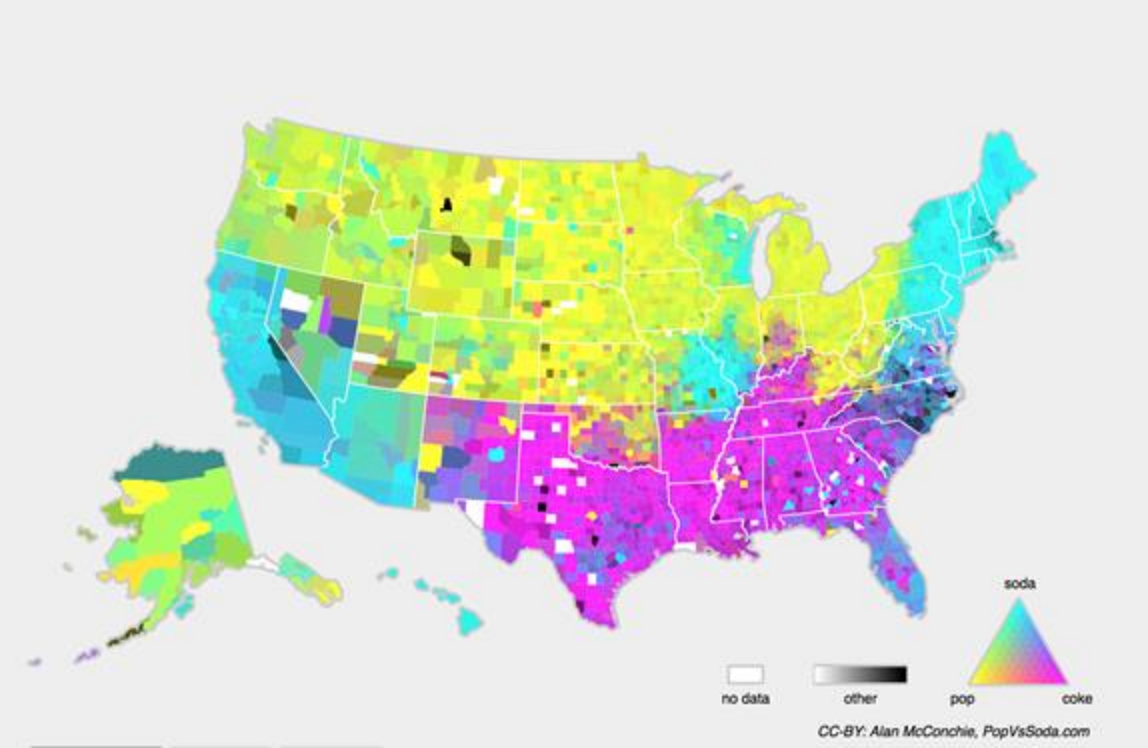
K-NN was trained using the default arguments in scikit-learn. ---> Default args in sklearn change!

Discuss Generalization

Many types of distribution shift:

- Covariate (feature) shift
- Label shift
- Concept shift

Comment on any generalization issues you foresee!



Example 1 (sparse on details)

The data used in this research was collected from a public dataset of images. The dataset contains 10,000 images, each of which is labeled with one of 10 different categories. The images were split into two sets: a training set and a test set. The training set was used to train the machine learning algorithm, and the test set was used to evaluate the performance of the algorithm.

The machine learning algorithm that was used in this research is a convolutional neural network (CNN). CNNs are a type of machine learning algorithm that is well-suited for image classification tasks. The CNN that was used in this research was trained using the Adam optimizer and the cross-entropy loss function.

The results of the research were evaluated using the accuracy metric. Accuracy is the percentage of images that the algorithm correctly classified. The algorithm achieved an accuracy of 95% on the test set.