2/26/2021
~~05:48~~

Hands On ML Ch. 1: ML Landscape

- Regression RF & SVM (RNN, CNN, or Transformer)

- Types
  - Supervision
  - Online (real-time) or incremental (batch)
  - Compare new data to known data or detect patterns & build predictive models (instance based vs. model-based)

(i) Supervised / Unsup/RL
desired solns = labels

set of features are predictors

∴ feature_list = predictors

predictors = {feature1, feature 2, etc.}

attr = value

feature

input features = predictors ∴ multiple output values

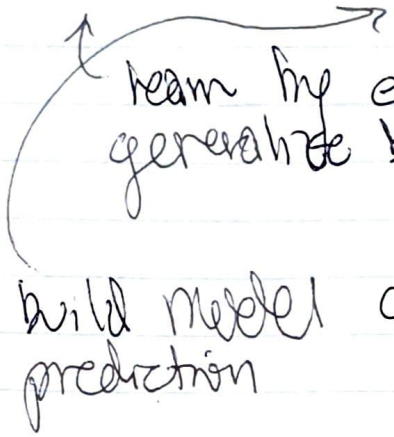
- Feature extraction is specific to dimensionality reduction

PRACTICAL TIP } Dim reduction before feeding to another ML algo

Pg 2

(ii) Batch & online
--all available data vs.
-- vs.
--train system incrementally in mini batches
note: online is really incremental learning + only loads part of what may be a huge dataset.

(iii) Instance vs. Model-Based



Performance measure?
- utility fxn ~~or or~~ for how good
- cost fxn for how bad

Question: Clustering of outliers? Load data, prepare data, plot data, select ~~linear~~ model, train it, test, ...

(iv) Features
--selection --extraction --~~regularization~~

(v) Overfitting: good performance on training but poor on generalization (validation)
--regularization (make model simpler)

(vi) Underfitting

pg 2

☑ Testing & Validation
--error rate on new cases → generalization error

More ~~underfitting~~ overfitting: IF training error is low & generalization is high

☑ Hyperparameter Tuning & model selection

Holdout validation → part of the training set is used to evaluate candidate ~~neo~~ models.

Cross validation → uses small validation sets