Machine Learning Exam Notes (1)

pill 1

▷ Machine Learning:

- Improve the performance of a software system, based on previous experience
- Set of methods that can automatically detect paterns in data, and then we the uncovered patterns to predict future data, or to perform other kinds of decisions making under uncertainty.

▷ Key elements:

- There is a pattern
- We cannot pin it down mathematically
- We have data on it

▷ Types of M.L.:

- Predictive/Supervised learning: given a labelled data set $D = \{(x_i, y_i)\}_{i=1}^{N}$ pairs called training set, find a mapping from x to y. (regression, classification)
- Descriptive/Unsupervised learning: given a data set $\{x_i\}_{i=1}^{N}$ find something interesting or useful about their structure. (density estimation, dimensionality reduction)
- Reinforcement learning: given an external system upon which you can exert control action a and receive percepts p, a reward signal r indicating good performance, find mapping from P->A that maximizes some long-term measure of r.

▷ Modeling a supervised ML problem

- If the answer to the question is YES/NO or a finite set of answers => Clasification
- If the question is a prediction of a quantity => Regression

▷ Evaluation of the model: we choose the model with the higher validation accuracy

↳ $D = D_{train} \cup D_{validation} \cup D_{test}$

pill 2

▷ Dealing with missing data: Check pill 2

- inputing: The process of replacing missing data with another value
- one-hot encoding: Dummy variables recode one feature into K-1 new feat.
- hashing trick: convert categorical features (mostly strings, IDs ...) into numerical values (technique of dimensionality reduction)
  - ↳ Hashing function: random & consistent

p113

▷ Confusion Matrix

- TP (True Positives): When the classifier predicts a sample as positive and it is positive
- FP (False Positive): When the classifier predicts a sample positive and it is not positive, its negative.
- TN (True Negative): When the classifier predicts a sample negative and it is negative
- FN (False Negative): When the classifier predicts a sample negative and it is not negative, it's positive

		Gold Standard Positive	Gold Standard Negative	
Prediction	Positive	TP	FP	→ Precision
Prediction	Negative	FN	TN	→ Negative Predictive Value
		↓ Sensitivity (Recall)	↓ Specificity	

- Accuracy:

$$\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

- Column-Wise

↳ Sensitivity or Recall $= \dfrac{TP}{\text{Real Positives}} = \dfrac{TP}{TP + FN}$

↳ Specificity $= \dfrac{TN}{\text{Real Negatives}} = \dfrac{TN}{TN + FP}$

- Row-Wise:

↳ Precision or Positive Predictive Value $= \dfrac{TP}{\text{Pred. Pos.}} = \dfrac{TP}{TP + FP}$

↳ NPV $= \dfrac{TN}{\text{Pred. Neg.}} = \dfrac{TN}{TN + FN}$

- $F1\text{-score} = 2 \cdot \frac{precision \cdot recall}{precision + recall}$

↓
unbalanced dataset

$$= 2 \cdot \frac{\frac{TP}{TP+FP} \cdot \frac{TP}{TP+FN}}{\frac{TP}{TP+FP} + \frac{TP}{TP+FN}} = \frac{2TP}{TP+FN+TP+FP}$$

- True Positive Rate = TPR = sensitivity (Recall) = $\frac{TP}{TP+FN}$
- False Positive Rate = FPR = 1 - TPR = $\frac{FP}{TN+FP}$
- ROC (Receiver Operating Characteristic): It shows how the TPR changes when the FPR changes

p.116

▷ Bias: As the number of training samples increase both errors ~~trends~~ (train error, test error) tent to the same value
or
The value in which both errors converge is called Bias

▷ Variance: The difference between bias & test error

▷ Overfitting: As the complexity increases the training error is reduced but above a certain complexity level the test error increases.

↳ To cure overfitting:
we want to control out of sample error: E_{out}
- Using cross-validation (simulate E_{out} & check out with unseen data)
- Regularization (minimize complexity of the model)
- Ensemble techniques

p.117

▷ Subgradient method:
- used for non-differentiable functions
- for solving large-scale opt. problems

↳ Stochastic gradients:
instead of using the exact gradient, these methods use a stochastic estimate of the gradient at each iteration
=> typically compute using subset (mini-batch) of the data
↳ introduces randomness & noise into the optimiz. process.

pill 8

Regularization: a technique that adjusts the learning objective / loss function to penalize the complexity of the model.
The primary goal is to prevent overfitting by adding a penalty term that discourages the model from fitting the training data too closely

Data augmentation: By creating artificially new data we can avoid overfitting.

[Overfitting occurs when the relationship between the complexity of the method and the number of available samples is large.]

-> Ways of doing that:

* Estimate the probability density function and then sample from that.

* Add noise to the current data while perserving the label from the original data

Dropout technique: Randomly deactivate a fraction of neurons during each forward and backward pass in training.

L1 regulation: Add penalty on the cost function and put the coefficients equal to zero.

-> Combined with linear models too

L2 regulation: Add penalty on the cost function and make the coefficients as ~~as~~ small as possible not exactly zero