CSE 446 Midterm Study Topics

Instructor: Sergey Levine

Summary

The purpose of this document is to provide a list of topics covered in the first half of the course. For the midterm exam, you will be expected to know each of these topics, as well as the corresponding derivations. It would also be a good idea to understand the relative strengths and weaknesses of each method we covered, you may be asked to choose a machine learning method for a given problem and justify your choice.

1 General machine learning concepts

- The four parts of a machine learning problem: data, hypothesis space, objective, algorithm.
- How we partition our data: training set, validation set, test set.
- Overfitting: definition, what causes it, how it can be mitigated.
- Bias and variance: what is it? What causes high variance? What causes bias? How does this relate to overfitting?
- Choosing hyperparameters: validation set, K-fold cross-validation, hold-one-out cross-validation.
- Basic probability: conditional distributions, Bayes rule.

2 Decision trees

- Why can't we simply use the space of all possible boolean functions?
- What is a decision tree?
- Why is it difficult to optimize decision trees exactly?
- Entropy, conditional entropy, and information gain.
- Why do decision trees overfit?
- Mitigating overfitting: pruning, fixed depth, Chi squared test, etc.
- How (and why) to use a validation set for pruning.

3 Probabilistic estimation

- Maximum likelihood estimation (MLE): what is it, what is objective.
- General idea of how to find the MLE solution for an arbitrary probability distribution.
- Bayes rule and how to compute a distribution over parameters.
- What is a prior (on parameters θ)?
- Some common choices of prior for distributions covered in lecture.
- Maximum a posteriori (MAP): what is it, what is objective.
- General idea of how to find the MAP solution for an arbitrary probability distribution with an arbitrary prior (assuming it can be done in closed form).

4 Linear regression

- Linear regression: data, hypothesis space, objective, algorithm.
- Probabilistic model corresponding to linear regression.
- Features $h(\mathbf{x})$: what are they, how are they used, some common choices (e.g. bias term).
- MLE vs MAP for linear regression: what does MLE look like? What does MAP look like (hint: ridge regression and LASSO).
- How to choose λ for ridge regression and LASSO, what its effect is on the bias and variance.
- Difference between ridge regression and LASSO: difference in objective, difference in effect on the weights $\mathbf{w}$.

5 Naïve Bayes

- Naïve Bayes: data, hypothesis space, objective, algorithm.
- Feature independence assumption: what does it mean? Why is it good? Why is it bad?
- Learning the conditional and prior distributions in Naïve Bayes (algorithm).
- MLE vs MAP in naïve Bayes, Laplace smoothing.
- Inference (how to obtain most likely $y^\star$ for a new input $x^\star$).
- Applications: bag of words model, continuous features.
- Bayesian networks: very basic understanding of what they are, how to interpret a graphical representation, how to estimate conditional proba-

bility distributions (same as in naïve Bayes); not required: inference in complex Bayesian networks, applications.

6 Logistic regression

- Logistic regression: data, hypothesis space, objective, algorithm.
- Conditional log-likelihood vs generative log-likelihood (in naïve Bayes).
- Computing derivatives with respect to $\mathbf{w}$ and using gradient ascent.
- MLE vs MAP in logistic regression: regularization, how to choose λ (hint: same idea as linear regression).
- Multiclass logistic regression.
- Comparison with naïve Bayes: which one overfits more? which one can handle more features? when to use one vs the other?

7 Neural networks

- Neural networks: data, hypothesis space, objective, algorithm.
- How do we get a neural network from logistic regression?
- Why does having more than one layer of weights (i.e. going beyond logistic regression) help?
- What are the parameters of a neural network? What are the hyperparameters?
- Backpropagation algorithm.
- Stochastic gradient descent (SGD).
- Multiclass classification: softmax function.
- Not on the midterm (this was presented just for your information): convolutional neural networks