CS 181 Midterm 1 Checklist

Due: March 1, 2022

Introduction

Below, we outline some key concepts and ideas you should be comfortable with for Midterm 1. Under each topic, we have divided this into three sections:

1. Items to know
2. Things to be able to work through when given information/formulae
3. Items out of scope

Note that this list is not exhaustive, but we hope it is illustrative. We encourage you to review the course textbook, section materials, such as section questions, homework materials, and the midterm practice problems for a full picture.

For emphasis: the midterm is not about memorization but will be designed to test your conceptual and analytical understanding.

1 Mathematical/Statistical Techniques

1.1 Items to know

- How to use linearity of expectation to compute expected square loss.
- Lagrange multipliers.
- Iterated chain rule when taking derivatives of elementary functions.
- Comfortability with performing operations on matrices and determining dimensions of matrix expressions.
- Familiarity with standard statistics techniques (Bayes Rule, MLE)

1.2 Things to be able to work through when given information/equations

- Given formulas/equations of equations composed of elementary functions, compute gradients with respect to vectors.
- Given PDFs and conjugacy equations, use these to provide expressions for the posterior.

1.3 Items out of scope

- Deriving conjugacies or using conjugacy statements from memory (e.g. Normal-Normal).

2 Regression

2.1 Items to know

- What is the "bias trick" for rolling the bias into $\mathbf{w}$?
- How does Kernelized regression work? What are its limitations?
- How does kNN work and what are its limitations?
- What is the least squares loss function?
- How is $\mathbf{w}^*$ derived in a linear regression problem?
- What is a basis function, and why would we use it?
- What is regularization, and when do we use it? How does the loss function change for Lasso and ridge regression?
- What is the bias variance tradeoff, and how does it relate to overfitting, underfitting, regularization, and increasing the size of the training set?
- What is a posterior distribution, posterior predictive distribution, and marginal likelihood?

2.2 Things to be able to work through when given information/equations

- You should be familiar with the mathematical techniques required to find $\mathbf{w}_{ridge}$. Sketch the derivation, and how is this similar to the derivation for $\mathbf{w}^*$ for OLS?
- Given a likelihood function (you do *not* need to remember PDFs), how is the MLE computed?
- Given a formula for kernel such as $K(x, x') = \exp\frac{-\|x-x'\|^2}{\tau}$, explain the effect of parameters such as τ for different use cases. Compare how this is similar or different to KNN.
- Given a Bayesian linear regression setup, how would the MAP be computed?
- Given expressions for conjugate distributions, and forms of posteriors, work through finding a posterior parameter or the posterior predictive for a given setting.

2.3 Items out of scope

- Second-order or other optimization methods (anything other than simple closed-form optimization and the gradient descent setup)

3 Classification

3.1 Items to know

- Role of hinge loss vs 0/1 loss vs logistic loss?
- What is the idea behind gradient descent (i.e. intuitively, why does it work)?
- What is the difference between gradient descent and stochastic gradient descent?
- If the data is linearly separable, are there any guarantees about perceptron's behavior? What is the intuition behind this?

- What can we say about the shape of the decision boundary of a logistic regression?
- Can you explain all the parts of the likelihood function for two-class and multi-class logistic regression?
- How do we work with generative models for classification? What is the Naive Bayes model?
- What is the difference between a discriminative model like Logistic Regression and a generative model like Naive Bayes?

3.2 Things to be able to work through when given information/equations

- Gradients for negated log likelihood of logistic regression
- For a generative model and give class priors and class-conditional distributions, work with the (full) log-likelihood (using Lagrangian method where needed)
- Given parametric as well as non-parametric classifiers (such as kNN) and a particular dataset, how do we expect different classifiers to perform?

3.3 Items out of scope

- Performing multiple, numerical steps of an iterative optimization such as gradient descent.
- Deriving basis functions to represent data in a space where it is separable (this is in scope for simple cases where such a transformation is easily discernible).

4 Neural Networks & Model Selection

4.1 Items to know

- Why do we need activation functions? What happens if we don't have them?
- What is backpropagation (theoretically), and why is it useful?
- How can neural networks be used for regression tasks and how can they be used for classification tasks?
- How can model selection methods such as cross-validation, and regularization be useful in supervised learning?
- What do we mean by the bias and variance of a model?
- What is the bias-variance tradeoff, and what are some remedies we have to resolve it?
- What is cross-validation and how does it work?

4.2 Things to be able to work through when given information/equations

- Given a particular task for a neural network, what is a good choice for a loss function (e.g., least squares vs softmax)?
- How do we go through a forward pass?
- Given a particular neural network architecture, how might changes to architecture (e.g. adding a layer, changing the structure) intuitively affect the model's performance?

- Given a simple neural network, determine whether the model is able to perform a particular classification task succesfully.
- Given a simple (not necessarily named) loss function for a neural network, deduce why it is intuitive for the task at hand.

4.3 Items out of scope

- Deriving backpropagation by hand.
- Extensions to convolutional or recurrent neural networks.