CS583A: Coverage of the Final Exam

April 30, 2019

1 Vector and matrix basics.

- Definitions of the ℓ_p vector norms for $p > 0$.
- Definitions and properties of the ℓ_p-norm balls.
- Matrix trace and matrix norms.
- Singular value decomposition (SVD) and truncated SVD.
- The BLAS and LAPACK libraries for matrix computation. (What are level 1, 2, and 3 BLAS?)

2 Scalar, vector, and matrix calculus.

- Derivative of a scalar w.r.t. a vector.
- Derivative of a vector w.r.t. a scalar.
- Derivative of a vector w.r.t. a vector.
- Derivative of a scalar w.r.t. a matrix.
- Chain rule.
- Always check the shape of a derivative!
- Subgradient and subdifferential.

3 Convex sets, convex functions, and convex optimization.

Convex sets.

- Definition of convex set.
- Typical examples of convex set and nonconvex set.

Convex functions.

- Definition of convex function.
- Definition of Hessian matrix.
- Definition of positive semi-definite.
- For convex function, the Hessian matrix is everywhere positive semi-definite.

Convex and nonconvex optimization.

- Definition of convex optimization.
- Definitions of objective function, constrains, and feasible set.
- For convex optimization, local optimum is global optimum.
- For convex optimization, the first-order optimality condition ($0 \in \partial f(\mathbf{w}^\star)$) implies $w^\star$ is a global minimum.
- For nonconvex optimization, there are saddle points.
- Definition of saddle points.
- For high-dimensional nonconvex optimization, #saddle points $\gg$ #local minima $\gg$ #global minima.

4 Machine Learning Basics

The four ML tasks.

- Definitions of regression, classification, clustering, and dimensionality reduction.
- Difference between regression and classification.
- Supervised learning and unsupervised learning.

Classification.

- Binary classification and multi-class classification methods, e.g., logistic regression, support vector machine (SVM), softmax classifier, and k-nearest neighbor (KNN).
- Linear classifiers include logistic regression, SVM, and softmax classifier.
- Nonlinear classifiers include KNN, kernel SVM, neural networks.
- What classification method is most suitable if #classes is millions?
- Standard evaluation metrics: accuracy, classification error rate, top 1 classification error, top 5 classification error.
- Evaluation metrics for class-imbalanced problems: true positive, true negative, false positive, false negative, ROC curve, precision, and recall.

Clustering.

- Clustering tasks are unsupervised learning.
- The k-means clustering method (a combinatorial optimization model).
- Lloyd's algorithm for approximately solving the k-means model.

Dimensionality reduction.

- Unsupervised learning methods: PCA and autoencoder.
- Supervised learning method: linear discriminant analysis.

Model capacity, overfitting, and underfitting.

- What controls model capacity? E.g., degree of polynomial in polynomial regression, number of layers and width of layers in neural networks, etc.
- What are overfitting and underfitting?
- How to alleviate overfitting and underfitting? More training samples, regularization, and data augmentation.

Hyper-parameters and cross-validation.

- Examples of hyper-parameters: degree of polynomials, regularization parameter, neural network structure, optimization algorithms. (The model itself is a hyper-parameter.)
- Training set, validation set, and test set.
- Never use the test for hyper-parameter tuning.

5 Convolutional Neural Networks

Convolutional operations

- Definitions of patch, filter, matrix and tensor convolution, zero-padding, stride, etc.
- Calculate the output shape given the input shape, filter shape, stride, and zero-padding.

Convolutional neural networks

- Using Keras to implement convolutional layer, pooling layer, flatten layer, and dense layers.
- Being able to choose appropriate activation functions.
- Given the input shape, filter number, filter shape, stride, zero-padding, and pool size, infer the **output shape** and **number of parameters**.

- Tricks for allevating overfitting: regularization, data augmentation, and pretrain (with the bottom layers frozen).
- Other tricks for improving the test error: multi-task learning and ensemble method.
- What is dropout? What is the best place to insert a dropout layer?
- How to properly use pretrain? How to use fine-tuning?
- What is feature scaling? What is batch normalization?

6 Recurrent Neural Networks

- Definitions of token, sequence, encoding, word embedding, and vocabulary.
- Using Keras to implement SimpleRNN layer, LSTM layer, and Embedding layer.
- Given the sequence length, vocabulary size, and embedding dimension (shape of x), infer the **output shape** and **number of parameters** in an embedding layer.
- Given the sequence length, embedding dimension (shape of x), and state dimension (shape of h), infer the **output shape** and **number of parameters** in a SimpleRNN or LSTM layer.
- The advantage of LSTM over RNN.
- How to make LSTM less likely to forget?
- How to alleviate overfitting?
- Definitions of Seq2Seq models, attention, self-attention, multi-head attention, and Transformer model.