MATH 270 SCHEDULE

TONY FENG

This schedule is subjective to change.

WEEK 1: INTRODUCTION TO NEURAL NETWORKS (SEP 8)

1.1. **Topics.** Supervised learning, regression, classification. Neural networks: activation functions, loss functions, gradient descent, backpropagation.

1.2. **Supplemental references.** Bishop §1 and §4 – §8.

WEEK 2: INFORMATION THEORY (SEP 15)

2.3. **Topics.** Shannon information and entropy. Conditional entropy, mutual information, KL divergence, Gibbs' inequality. Compression, arithmetic coding, relation to generative modeling.

2.4. **Supplemental references.** McKay, §1–6.

WEEK 3: STATISTICAL INFERENCE (SEP 22)

3.5. **Topics.** Central Limit Theorem. Estimators: Maximum Likelihood Estimation, cross-entropy, Bayesian inference. Evaluating estimators: MSE and MAE.

3.6. **Supplemental references.** Berger–Casella, §7.

WEEK 4: OPTIMIZATION (SEP 29)

4.7. **Topics.** Optimizers: momentum, RMSProp, Adam. Normalization: Kaiming normalization, Batch normalization, Layer normalization. Regularization: weight decay, dropout. Adversarial attacks.

4.8. **Supplemental references.** Bishop, §9. Kaplan, §4.

WEEK 5: CONVOLUTIONAL NEURAL NETWORKS (OCT 6)

5.9. **Topics.** Inductive bias: locality, equivariance and invariance. CNN architecture: filters, feature maps, pooling.

5.10. **Case study.** ImageNet Challenge. AlexNet, VGGNet, ResNet, residual learning, interpretability.

5.11. **Supplemental references.** Bishop, §10.

WEEK 6: RECURRENT NEURAL NETWORKS (OCT 13)

5.12. **Topics.** RNN Architecture: hidden states, unrolling, memory. Representation learning. Encoder-decoder structure. Sampling: greedy, beam search, temperature sampling.

5.13. **Case study.** Natural language processing: classification and machine translation.

    (1) Cho, et. al, *Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation.* (2014).
    (2) Sutskever et. al, *Sequence to Sequence Learning with Neural Networks.* (2014).
    (3) Bahdanau et. al, *Neural Machine Translation by Jointly Learning to Align and Translate.* ICLR (2015)

5.14. **Supplemental references.** geeksforgeeks and Andrej Karpathy's blog post.

Week 7: Transformers (Oct 20)

5.15. **Topics.** Attention mechanism. Transformers.

5.16. **Case studies.** BERT, Vision Transformers, Transformer-XL.

5.17. **Supplemental references.** Vaswani et. al, *Attention is all you need* and Harvard NLP group's reproduction. Bishop §12.

Week 8: Large Language Models (Oct 27)

5.18. **Topics.** Decoder-only LLM architecture. Generative pretraining. Supervised fine-tuning. Data mixture, filtering, and deduplication. Chinchilla scaling laws.

5.19. **Case studies.** GPT-1, GPT-2, GPT-3. LLAMA models. Scaling laws.

5.20. **Supplemental references.**

Week 9: Generative Adversarial Networks and Variational Autoencoders (Nov 3)

5.21. **Topics.** Generative Adversarial Networks. Variational Autoencoders. Evidence Lower Bound. Reparametrization trick.

5.22. **Case study.** Image generation. Cycle-GAN, β-VAE.

5.23. **Supplemental references.** Bishop, §17 and §19.

Week 10: Diffusion Models (Nov 10)

5.24. **Topics.** Diffusion models. More Evidence Lower Bound.

5.25. **Case study.** More image generation. Stable diffusion.

5.26. **Supplemental references.** Bishop, §19 and §20.

Week 11: No class (Nov 17)

Week 12: Reinforcement learning I: value function methods (Nov 24)

5.27. **Topics.** Markov decision processes. Prediction and Control: value function optimization and policy optimization. Monte Carlo methods, temporal-difference learning, SARSA, value function approximation.

5.28. **Case study.** Game-playing: Atari, AlphaGo, AlphaZero.

5.29. **Supplemental references.** Sutton–Barto, §3–5. David Silver's Youtube course.

Week 13: Reinforcement learning II: policy optimization (Dec 1)

5.30. **Topics.** Proximal policy optimization.

5.31. **Case study.** Reasoning: DeepSeek-R1, AlphaGeometry, AlphaCode.

5.32. **Supplemental references.** Denny Zhou's Youtube talk and slides.