# Prerequisite Expectations for CPSC 540-542

CPSC 540, 541, and 542 are Graduate level courses, which require a certain amount of prerequisite/foundational knowledge so that we can build up to more advanced topics. Below is a rough list of major topics that we expect students to be familiar with (though it's not guaranteed to be exhaustive). Links to resources are also provided for students who need to review or learn this information.

## Topics

### Probability

- **Probability Density Functions**
- **Common Distributions (Normal, Uniform, Bernoulli)**
- **Probabilities/Conditional Probabilities**
- **Expected Values, Variance and Covariance of Combinations of Random Variables**
- **Bayes' Theorem**
- Maximum Likelihood Estimation

### Linear Algebra

- Matrix Multiplication
- Eigendecomposition
- **Linear Combinations**
- **Vector Spaces**

### Calculus

- Derivatives
- Integrals
- Partial Derivatives/Gradients
- Series (Taylor Series, Maclaurin Series)
- **Matrix Calculus

### Information Theory

- **Entropy**
- **Cross Entropy**
- **Kullback-Leibler (KL) Divergence**
- **Shannon-Jensen (SJ) Divergence**
- **Wasserstein Distance**

### Machine Learning

- Bias-Variance Tradeoff
- Regularization

- Feature Engineering
- Performance Metrics (e.g. *MSE*, *MAE*, *R2*, *Accuracy*, *Precision*, *Recall*, *ROC AUC*, *Calibration*)
- Supervised vs. Unsupervised Models
- Clustering (*K-Means*, *Gaussian Mixture Models*, *DBSCAN*, *Hierarchical Clustering*)
- Dimensionality Reduction
- Classification vs. Regression Models
- Tree Based Models
- Ensemble Models
- Linear Regression
- Logistic Regression
- K-Nearest Neighbors
- Naive Bayes

## Deep Learning

- Loss Functions (including common ones like cross-entropy, mean squared error...etc)
- Gradient Descent
- Backpropagation
- Computer Vision (e.g. Convolutional Layers, Transposed Convolutional Layers, Pooling, Strides, Weight Sharing, Translational Invariance, GradCAM, Depthwise Separable Convolutions)
- Autoencoders
- Sequential Models (e.g. Recurrent NNs, LSTMs, GRUs, Transformers, Text Processing, Word Embeddings)
- Generative Models (e.g. Density Estimation vs. Sample Generation, GANs, VAEs)

# Sample Questions (Test Yourself)

1. Explain the role of Backpropagation in training a Deep Neural Network.
2. Explain how Dropout reduces the likelihood of overfitting in a Neural Network.
3. How are Generative Density Estimation Models different from Generative Sample Generation models? Give at least one example of each.
4. Describe the goal of regularization, and give at least three examples of tools/techniques we use to regularize Machine Learning and/or Deep Learning Models.
5. How do overcomplete Autoencoders learn a useful representation, even though their hidden representation dimension is *larger* than the input/output dimension?
6. Explain why Calibration is important for classification models.
7. Explain at least one difference and one similarity between RMSProp and AdaGrad.
8. What is one thing Random Forests do to prevent overfitting?

9. Explain the difference between Random Forests, and Gradient Boosting Trees.
10. What is the purpose of a loss function? Give two examples of common loss functions.
11. What is “Naive” about the Naive Bayes algorithm?
12. Explain at least two differences between the assumptions Gaussian Mixture Models make about clusters, compared to K-Means.
13. Explain how Polynomial Regression allows you to make non-linear predictions using a linear model.
14. What does the ROC AUC value tell you about a classification model?
15. Explain how Principal Component Analysis does Dimensionality Reduction without doing Variable Selection.
16. Why do we want leaf nodes in a Decision Tree to have low Entropy?
17. Explain how Maximum Likelihood Estimation helps us choose Logistic Regression coefficients.
18. What is Translational Invariance (in Convolutional NNs)? Describe two things (either in the architecture or training) that encourages CNNs to have translational invariance.
19. Explain the Gradient Descent Update rule.
20. What is the difference between Convolutional Layers and Transposed Convolutional Layers?
21. When would we want to use Max Pooling, vs Strides to downsample images in a CNN?
22. Explain the vanishing gradient problem.
23. What is z-scoring, and why is it useful (give at least 2 reasons).
24. Explain Attention (in the context of Transformers).
25. What is positional encoding, and how was it implemented in the 2017 paper “Attention is All You Need”?
26. How are Variational Autoencoders different from Vanilla Autoencoders?
27. Describe how a Wasserstein GAN is different from a regular GAN, and explain why these changes are beneficial.
28. How do models like word2vec create word embeddings?
29. What’s the difference between the two probabilities P(dog) and P(dog | over 20)?
30. What is the difference between Self-Attention and Cross-Attention?
31. Explain how Queries, Keys, and Values play a role in Self-Attention (in a Transformer architecture).
32.

# Resources

## Probability

- 📚Berger and Casella Chapters 1,2,3, and 4
- 📚Mathematics for Machine Learning Chapters 6

## Linear Algebra

- 📚Mathematics for Machine Learning Chapters 2 and 4
- 🖥️3Blue1Brown Linear Algebra Series

## Calculus

- 📚Mathematics for Machine Learning Chapter 5
- 🖥️3Blue1Brown Calculus Series
- 📄Khan Academy: Lagrange Multipliers and Constrained Optimization
- 📄Common Taylor Series
- **Matrix Calculus

## Information Theory

- Entropy
- Cross Entropy
- Kullback-Leibler (KL) Divergence
- Shannon-Jensen (SJ) Divergence
- Wasserstein Distance
- 📄Count Bayesie: KL Divergence
- https://web.stanford.edu/class/stats311/lecture-notes.pdf
- http://madhu.seas.harvard.edu/courses/Spring2019/
- https://www.inference.org.uk/itprnn/book.pdf

## Machine Learning

- 🖥️Dr. Parlett's 392 Lecture Videos
- 📚Introduction to Statistical Learning with Applications in R (ISLR)
- 📚The Elements of Statistical Learning (ESL)
- 📄https://github.com/afshinea/stanford-cs-229-machine-learning/blob/master/en/super-cheatsheet-machine-learning.pdf

## Deep Learning

- Deep Learning (Multiple Topics)
    - 🖥️Dr. Parlett's 393 Lecture Videos
    - 🖥️3Blue1Brown Neural Network Series
    - 📚Deep Learning with PyTorch
    - 📚Deep Learning
    - 📚Dive into Deep Learning
    - 📄https://github.com/afshinea/stanford-cs-229-machine-learning/blob/master/en/cheatsheet-deep-learning.pdf

- Math of ML
  - 🖥️StatQuest Backpropagation Main Ideas
  - 🖥️StatQuest Backpropagation Details Pt 1:
  - 🖥️StatQuest Backpropagation Details Pt 2:
  - 📄Ruder.io: An Overview of Gradient Descent Optimization Algorithms
  - 🖥️Code Emporium: Gradient Descent, the Math You Should Know
  - 📄Jacobian vs. Hessian vs. Gradient
- Convolutional/Vision Neural Networks
  - 📄Backpropagation In Convolutional Neural Networks
  - 🖥️Code Emporium: Convolutional Neural Networks Explained!
- Autoencoders
  - 📄Jeremy Jordan: Autoencoders
  - 📄Geeks for Geeks: Contractive Autoencoders
- Recurrent Neural Networks (RNN, LSTM, GRU)
  - 🖥️StatQuest LSTM's Clearly Explained
  - 🖥️Code Emporium: LSTM Networks Explained!
- Transformers
  - 📄e2ml: Transfomers from Scratch
  - 📄The Annotated Transformer
  - 📄The Illustrated Transformer
- Generative Models
  - 🖥️Code Emporium: GANs
  - 📄Jeremy Jordan: Variational Autoencoders