

Study Guide

Last modified: 2026-10-28

Below are specific criteria you can use to focus your studying. They cover many of the major points, but they *are not guaranteed to be exhaustive*.

Each module lists its outcomes, followed by things you should be able to do to check your understanding. The modules are numbered and ordered as the course's Canvas modules are, and cover what the calendar's lectures cover. Outcomes about topics that the calendar lists as optional reading are optional too, and the last group of modules, on optional reading, is entirely optional.

M0: Prerequisites

Outcomes

- A basic understanding of gradients
- An understanding of linear functions and inner products
- A basic understanding of hyperplanes and half spaces
- A basic understanding of joint, marginal and conditional probabilities¹
- A basic understanding of the law of total probability², the chain rule of probability³, and Bayes' rule⁴
- A basic understanding of independent and conditionally independent variables⁵

To check your understanding, you should be able to...

- Compute simple gradients
- Identify the shapes of gradients
- Reason about the relative inner products between pairs of vectors
- Determine if two probability expressions are greater, less than, or equal
- Demonstrate the equivalence of two probability expressions
- Relate the dimensions of hyperplanes to the two halfspaces they divide space into
- Demonstrate the use of the law of total probability, the chain rule of probability, and Bayes' rule

M1: Course Overview

Outcomes

- An understanding of what machine learning is, and of learning from experience E at tasks T as measured by P
- An understanding of why some tasks call for learning rules from data rather than writing them by hand
- A basic understanding of how machine learning relates to artificial intelligence and deep learning and to statistics
- An understanding of the difference between prediction and inference

¹<https://morrison-lab.github.io/pds/random-variables.html#def-joint-distribution>

²<https://morrison-lab.github.io/pds/probability-basics.html#thm-total-prob>

³<https://morrison-lab.github.io/pds/probability-basics.html#thm-law-conditional-prob>

⁴<https://morrison-lab.github.io/pds/probability-basics.html#thm-bayes>

⁵<https://morrison-lab.github.io/pds/independence.html#def-indpt>

- An understanding of the parts of a learning problem: training data, a learning algorithm, training and the model it produces
- A basic understanding of structured and unstructured data

To check your understanding, you should be able to...

- Name the experience E , task T and performance measure P of a learning problem from its description.
- Explain when learning a rule from data is a better choice than writing it by hand.
- Classify a question about data as prediction or inference.
- Describe an input as a feature vector.
- Compute the sample mean squared error (sMSE) of a fitted model on a few examples by hand.

M2: Linear Regression

Outcomes

- An understanding of the regression problem

To check your understanding, you should be able to...

- Explain what it means for a function to be convex, and the implications this has for optimization
- State the equation for $h(x)$ for linear regression, and identify and count its model parameters
- State the equation for sample mean squared error (sMSE), and reason about its properties

M3: ML Fundamentals

Outcomes

- An understanding of the overall machine learning framework, and the prediction function $h(x)$ in particular
- An understanding of loss, risk and empirical risk
- A basic understanding of the kinds of machine learning problem (e.g. clustering, classification, regression, ranking, sequence labeling, structured prediction, etc.)
- A basic understanding of input features and feature vectors, including “one-hot” representations and normalization techniques
- An understanding of the issues of overfitting, underfitting, generalization, regularization, estimation⁶ bias and variance, and residual variance
- A basic understanding of training scenarios: supervised, unsupervised and semi-supervised
- A basic understanding of the distinction between probabilistic and non-probabilistic models, and between generative and discriminative ones
- An understanding of the role of training, validation and test sets

To check your understanding, you should be able to...

- Predict which of two models is more likely to overfit or underfit.
- Define key characteristics (e.g. supervised, probabilistic).
- Identify key characteristics from a problem description.
- Explain the role of the train, validation and test sets.
- Compare the expected empirical risk vs true risk for two models.
- Represent categorical variables as one-hot vectors.

⁶<https://morrison-lab.github.io/sds/estimation.html#def-estimate>

M4: Classification and Logistic Regression

Outcomes

- An understanding of the classification problem
- A basic understanding of decision boundaries and decision regions
- An understanding of the logistic sigmoid
- An understanding of maximum likelihood estimation and of the cross entropy loss it gives logistic regression
- An understanding of Newton's method and of when the coefficients diverge
- An understanding of gradients and how they relate to optimization

To check your understanding, you should be able to...

- Show how the shape of the logistic regression model changes depending on the parameters
- Sketch the decision boundary and decision regions for a logistic regression classifier
- Reason about the relative probabilities a logistic regression model would assign to different points
- Explain the gradient descent algorithm, including what role the gradients play, for training a linear model
- State the equation for $h(x)$ for logistic regression, and identify and count its model parameters
- State the equation for cross entropy loss, and reason about its properties

Optional: more than two classes

The calendar lists more than two classes as optional reading, apart from the softmax, which the neural network lectures use.

- An understanding of the softmax function
- Show how the softmax function transforms vectors (know the definition, but also show qualitatively how different vectors are transformed by softmax)
- Explain how the probabilities given by a multinomial logistic regression model relate to the model parameters
- State the equation for $h(x)$ for multinomial logistic regression, and identify and count its model parameters

M5: Performance Metrics

Outcomes

- An understanding of common performance metrics for classification, detection and regression

To check your understanding, you should be able to...

- Identify metrics with certain properties (e.g. differentiability)
- Identify metrics for certain tasks (e.g. detection)
- Compute and explain confusion matrices (multiclass and binary)
- Create binary detection confusion matrices from multiclass confusion matrices
- Describe how the metrics typically relate to each other (e.g. monotonic, inversely proportional)
- Anticipate the effects of changing the detection threshold on true and false positives and negatives
- Anticipate the effects of increasing or decreasing true and false positives and negatives when keeping all other values constant
- Identify the circumstances and model behaviors that lead to high or low performance under the various metrics
- Compute metrics numerically in simple cases

M6: Regularization and Model Selection

Outcomes

- An understanding of why we regularize
- An understanding of ridge regression and the lasso
- An understanding of hyperparameters and why training risk cannot choose the model
- A basic understanding of hyperparameter search strategies

To check your understanding, you should be able to...

- Explain how a regularization penalty trades training risk against model complexity.
- Compare the effects of the ridge and lasso penalties on a model's coefficients.
- Explain how to tune a hyperparameter by cross-validation.

M7: Support Vector Machines and Kernel Methods

Outcomes

- A thorough understanding of optimal hyperplanes
- An understanding of the concept of margin
- A basic understanding of the primal and dual forms of support vector machine (SVM) training
- An understanding of soft-margin classifiers
- An understanding of kernels and how they allow non-linear classification
- An understanding of the common ways to perform multiclass classification with SVMs

To check your understanding, you should be able to...

- Define the concept of the margin of a hyperplane, and explain why we might want to maximize it.
- Visually identify the margin of a hyperplane, and sketch the hyperplane with maximum margin.
- Identify when adding or removing training data points will change the optimal SVM hyperplane.
- Define and visually identify support vectors.
- Given some information about a problem, explain whether a support vector machine is a good fit for it.
- Describe what slack variables are and why they are needed.
- Explain the relationship between slack variables and the margin.
- Visually identify the slack variable values given points, a hyperplane and its margin.
- Describe how the hyper-parameter C (in C -SVM) relates to the margin.
- Describe how ν (in ν -SVM) relates to the margin and the number of support vectors.
- Explain what a kernel is, and how the “kernel trick” allows SVMs to do non-linear classification.
- Mathematically define the polynomial kernel and the Gaussian radial basis function kernel. Explain how the hyper-parameter for the polynomial kernel will affect the bias and variance.
- Reason about kernels as notions of similarity between points, and compare similarities under different kernels.
- Explain how, when and why you might use the parametric or non-parametric versions of SVMs.
- Explain how one-vs-one and one-vs-rest accomplish multiclass classification, and what the pros and cons of those two techniques are.

M8: Neural Networks and Backpropagation

Outcomes

- A basic understanding of polynomial feature transformation (and its limitations)
- An understanding of how neural networks relate to logistic regression
- An understanding of neural network architecture (layers, weights, diagrams, hyper-parameters, etc.)
- A thorough understanding of how neural networks can compute basic logical operations
- A basic understanding of the theoretical representational power of neural networks
- A basic understanding of neural network training (back propagation and stochastic gradients)

- An understanding of the activation and loss functions appropriate for classification and regression
- A basic understanding of the effect of minibatch size on stochastic gradient descent (SGD)

To check your understanding, you should be able to...

- List the three major approaches to doing non-linear classification, give an example classifier that works with each strategy, and list one pro and one con for each strategy.
- Sketch to illustrate how explicit basis expansion followed by linear classification can yield non-linear classification in the original input space.
- Define the neural network model as an equation relating the input to output.
- Explain how a neural network can be interpreted as a learned feature transformation followed by a linear model.
- Explain what aspects of a neural network might differ when the network is being used for classification vs regression.
- List some of the common non-linear activation functions and describe their properties.
- Explain what a “bottleneck” layer is and when it might be used.
- Design logistic nodes to implement specific logical functions, and be able to identify when a logical function can or cannot be implemented with a linear model.
- Describe the representational power of neural networks: what functions can they approximate?
- Explain why deep models may be superior to shallow ones for certain problems.
- Explain what “backpropagation” is and how it is used to train a neural network.
- Explain why deep models may be more difficult to train than shallow ones.
- Explain how the behavior of stochastic gradient descent might change as you vary the minibatch size.

M9: ML Experimentation and Ethics

Outcomes

- An understanding of the ML experimentation terminology we have adopted in class
- A basic understanding of data concepts such as benchmarks and synthetic data
- An understanding of the issues related to splitting data into train, validation and test sets
- An understanding of considerations for normalizing data, including cyclical encodings
- An understanding of the role of baselines, common baseline strategies, and considerations for selecting baselines
- A basic understanding of hyperparameter search strategies
- An understanding of the training loop, including order of operations and early stopping
- A basic understanding of results and analyses, including ablation studies

M10: Clustering

Outcomes

- An understanding of the clustering problem
- An understanding of unsupervised learning
- A thorough understanding of the k -means algorithm
- A thorough understanding of agglomerative clustering
- A thorough understanding of multivariate normal distributions
- A basic understanding of mixture models
- An understanding of Gaussian mixture models (GMMs)

To check your understanding, you should be able to...

- Analyze and anticipate the clustering behavior of the k -means algorithm.
- Analyze and anticipate the clustering behavior of agglomerative clustering with different cluster distance metrics (single link, complete link, etc.).

- Analyze and anticipate the shape and properties of a multivariate normal distribution as a function of the mean and covariance matrix.
- Identify, define, and describe the role and properties of the various parameters in a GMM.
- Compare Mahalanobis distances between points, and explain how that relates to multivariate normal distributions.
- Explain the objective and closed-form solution to the maximum likelihood estimation problem.
- Identify cases where mixture models would be better suited to density estimation than single Gaussians.

Optional reading

The calendar’s lectures skip the modules below, so everything in them is optional. The calendar’s optional reading list names the skipped parts of the other chapters.

O1: Machine Learning Theory

Outcomes

- A basic understanding of the Bayes decision rule and Bayes risk
- A basic understanding of zero-one loss and Bayes error
- A basic understanding of the curse of dimensionality
- A basic understanding of the No Free Lunch Theorem
- A basic understanding of the Ugly Duckling Theorem

To check your understanding, you should be able to...

- Define the Bayes decision rule and Bayes risk.
- Explain how Bayes risk informs our expectation of classifier performance.
- Explain how Bayes risk relates to Bayes error (including when and why they might differ).
- Explain how zero-one loss relates to Bayes error, and when you might prefer to use zero-one loss.
- Illustrate reducible error, Bayes error and residual variance visually.
- Explain the “loophole” that allows us to reduce the Bayes error and the residual variance: each is irreducible only given the inputs.
- Explain the consequences of the curse of dimensionality, and how it affects machine learning.
- Explain what the No Free Lunch Theorem means for machine learning.
- Explain what the Ugly Duckling Theorem means for machine learning.

O2: Non-Parametric Models and Ensembles

O3: Further Methods

Project ideas

Any technique from the notes can be tried on a real data set. Each data set below has a copy on this site that you can download, and a prediction question to start from.

Data	Source	Rows	Columns	Question to start from
Auto ⁷	UCI Auto MPG [1], from the StatLib library; this copy is the one in the ISLP and ISLR2 packages [2]	392	9	Predict a car’s fuel economy (mpg) from its engine and weight.
Boston ⁸	Harrison and Rubinfeld’s study of housing values in Boston suburbs [3]; the copy in ISLP and ISLR2 [2]	506	13	Predict a suburb’s median home value (medv) and ask which predictors matter.

⁷<https://morrisson-lab.github.io/lds/chapters/data/auto.csv>

⁸<https://morrisson-lab.github.io/lds/chapters/data/boston.csv>

Data Source	Rows	Columns	Question to start from
Carseats ⁹ : Simulated data set of child car seat sales, from the ISLR2 package [2]	400	11	Predict sales (Sales) at each store.
Hitters ¹⁰ : Major League Baseball statistics from the 1986 and 1987 seasons, from the ISLR2 package [2]	322	20	Predict a player’s salary (Salary) from many correlated statistics.
SmarterMarket ¹¹ : Daily returns of the S&P 500 stock index, from the ISLR2 package [2]	1250	9	Predict whether the market goes up or down (Direction) from the previous days’ returns.
Spam-UCI Spambase [4] base ¹²	4601	58	Classify an email as spam or not from word and character frequencies (Class).
WCGS ¹³ : The Western Collaborative Group Study of coronary heart disease [5]	3154	5	Predict coronary heart disease (chd) from age, body mass index, blood pressure and behavior pattern.
Optdigits ¹⁴ : UCI Optical Recognition of Handwritten Digits [6]; this file is its test set	1797	65	Classify a handwritten digit from an 8 by 8 grid of pixel counts.

The **Auto** file has fewer rows than the UCI original, which also lists the cars with a missing horsepower; the ISLP copy drops them. The **Optdigits** file is the UCI test set, with the digit label in its last column.

To go beyond these, search the official repositories of data sets:

- Kaggle Datasets¹⁵ hosts data sets from many fields, with competitions on some.
- The UCI Machine Learning Repository¹⁶ holds classic benchmark data sets, each with its own citation.
- OpenML¹⁷ lists data sets with standard train and test splits for comparing methods.
- Hugging Face Datasets¹⁸ holds text, image and audio data sets.

Before you pick a data set, read where it came from, and cite it the way its page asks.

- [1] R. Quinlan, “Auto MPG.” UCI Machine Learning Repository, 1993. doi: 10.24432/C5859H¹⁹.
- [2] G. James, D. Witten, T. Hastie, R. Tibshirani, and J. Taylor, *An introduction to statistical learning: With applications in Python*. Springer, 2023. doi: 10.1007/978-3-031-38747-0²⁰.
- [3] D. Harrison and D. L. Rubinfeld, “Hedonic housing prices and the demand for clean air,” *Journal of Environmental Economics and Management*, vol. 5, no. 1, pp. 81–102, 1978, doi: 10.1016/0095-0696(78)90006-2²¹.
- [4] M. Hopkins, E. Reeber, G. Forman, and J. Suermondt, “Spambase.” UCI Machine Learning Repository, 1999. doi: 10.24432/C53G6X²².
- [5] R. H. Rosenman, R. J. Brand, C. D. Jenkins, M. Friedman, R. Straus, and M. Wurm, “Coronary heart disease in the Western Collaborative Group Study: Final follow-up experience of 8 1/2 years,” *JAMA*, vol. 233, no. 8, pp. 872–877, 1975, doi: 10.1001/jama.1975.03260080034016²³.
- [6] E. Alpaydin and C. Kaynak, “Optical recognition of handwritten digits.” UCI Machine Learning Repository, 1998. doi: 10.24432/C50P49²⁴.

⁹<https://morrison-lab.github.io/lds/chapters/data/carseats.csv>

¹⁰<https://morrison-lab.github.io/lds/chapters/data/hitters.csv>

¹¹<https://morrison-lab.github.io/lds/chapters/data/smarket.csv>

¹²<https://morrison-lab.github.io/lds/chapters/data/spambase.csv>

¹³<https://morrison-lab.github.io/lds/chapters/data/wcgs.csv>

¹⁴<https://morrison-lab.github.io/lds/chapters/data/optdigits-test.csv>

¹⁵<https://www.kaggle.com/datasets>

¹⁶<https://archive.ics.uci.edu/>

¹⁷<https://www.openml.org/>

¹⁸<https://huggingface.co/datasets>

¹⁹<https://doi.org/10.24432/C5859H>

²⁰<https://doi.org/10.1007/978-3-031-38747-0>

²¹[https://doi.org/10.1016/0095-0696\(78\)90006-2](https://doi.org/10.1016/0095-0696(78)90006-2)

²²<https://doi.org/10.24432/C53G6X>

²³<https://doi.org/10.1001/jama.1975.03260080034016>

²⁴<https://doi.org/10.24432/C50P49>