1. Andrew Wu’s Machine Learning;
2. ---------------------------------------Chapter 1----------------------------------------
3. Ep 1-1: welcome to machine learning; (2019-6-30)
4. Ep 1-2: welcome; machine learning; (2019-6-30)
5. Ep 1-3: what is machine learning; supervised learning, unsupervised learning, reinforcement learning, recommender systems; (2019-6-30)

Reading: Tom Mitchell provides a more modern definition: "A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E." (2019-3-23)

1. Ep 1-4: supervised learning; examples; house price prediction; (2018-12-24)

Reading: In supervised learning, we are given a data set and already know what our correct output should look like, having the idea that there is a relationship between the input and the output;

Supervised learning problems are categorized into "regression" and "classification" problems; (2019-3-23)

1. Ep 1-5: unsupervised learning; cocktail algorithm; (2019-3-22)

Reading: We can derive this structure by clustering the data based on relationships among the variables in the data; (2019-3-23)

1. Ep 1-6: model representation; housing prices; training set of housing prices; (2019-6-30)

Reading: regression problem; classification problem; (2019-3-22)

1. Ep 1-7: cost function; hypothesis; cost function; (2018-12-26)

Reading: This function is otherwise called the "Squared error function", or "Mean squared error";

1. Ep 1-8: cost function 1: simplified example; (2019-3-18)

Reading: The best possible line will be such so that the average squared vertical distances of the scattered points from the line will be the least; (2019-3-23)

1. Ep 1-9: cost function 2: (2019-3-18)

Reading: A contour plot is a graph that contains many contour lines; (2019-3-23)

1. Ep 1-10: gradient descent; gradient descent algorithm; correct/incorrect; (2019-3-18)

Reading: The way we do this is by taking the derivative (the tangential line to a function) of our cost function; (2019-3-23)

1. Ep 1-11: gradient descent intuition; gradient descent algorithm; (2019-3-18)

Reading: when the slope is negative, the value of theta, start subscript, 1, end subscript θ1\theta\_1θ1​ increases and when it is positive, the value of theta, start subscript, 1, end subscript θ1\theta\_1θ1​ decreases; (2019-3-23)

1. Ep 1-12: gradient descent for linear regression; linear regression model; batch gradient descent; (2019-3-19)

Reading: When specifically applied to the case of linear regression, a new form of the gradient descent equation can be derived; (2019-3-23)

1. Ep 1-13: matrices and vectors; matrix, rectangular array of numbers; matrix element; (2019-3-19)
2. Ep 1-14: additional and scalar multiplication; matrix addition; scalar multiplication; (2019-3-19)
3. Ep 1-15: matrix vector multiplication; example; (2019-3-19)
4. Ep 1-16: matrix-matrix multiplication; having 3 competing hypotheses; (2019-3-19)
5. Ep 1-17: matrix multiplication properties; identity matrix; (2019-3-19)
6. Ep 1-18: inverse and transpose; matrix inverse; matrix transpose; (2019-3-19)
7. ---------------------------------------Chapter 2----------------------------------------
8. Ep 2-1: multiple features; hypothesis; (2019-3-19)

Reading: Linear regression with multiple variables is also known as "multivariate linear regression"; (2019-3-23)

1. Ep 2-2: gradient descent for multiple variables; hypothesis, parameter, cost function, gradient descent; (2019-3-19)

Reading: The gradient descent equation itself is generally the same form; we just have to repeat it for our 'n' features; (2019-3-23)

1. Ep 2-3: linear regression with multiple variables; feature scaling; mean normalization; (2019-3-19)

Reading: Where μi is the average of all the values for feature (i) and s, start subscript, i, end subscript sis\_isi​ is the range of values (max - min), or s, start subscript, i, end subscript sis\_is is the standard deviation; (2019-3-23)

1. Ep 2-4: gradient descent in practice II, learning rate; make sure gradient descent is working correctly; summary, alpha; (2019-3-19)

Reading: If alpha α\alphaα is too small, slow convergence; if alpha α\alphaα is too large: ￼may not decrease on every iteration and thus may not converge; (2019-3-24)

1. Ep 2-5: features and polynomial regression; housing prices prediction; polynomial regression; choice of features; (2019-3-19)

Reading: We can change the behavior or curve of our hypothesis function by making it a quadratic, cubic or square root function (or any other form); (2019-3-24)

1. Ep 2-6: normal equation; intuition; examples; m training examples, n features; gradient descent; normal equation; (2019-3-19)

Reading: This allows us to find the optimum theta without iteration; (2019-3-24)

1. Ep 2-7: normal equation and non-invertibility; normal equation; too many features; (2019-3-19)

Reading: Solutions to the above problems include deleting a feature that is linearly dependent with another or deleting one or more features when there are too many features; (2019-3-24)

1. Ep 2-8: working on and submitting programming; (2019-3-19)
2. Ep 2-9: basic operations; first use octave, then use C++ and java; (2019-3-21)
3. Ep 2-10: moving data around; size; length; load; who; save; (2019-3-21)
4. Ep 2-11: computing on data; transpose; magic(3); ceil, floor; dot product; (2019-3-21)
5. Ep 2-12: plotting data; plot(t,y1); xlable, ylabel; figure(1), figure(2); imagesc; (2019-3-24)
6. Ep 2-13: for, while, if; (2019-3-24)
7. Ep 2-14: vectorization; example; (2019-3-24)
8. ---------------------------------------Chapter 3----------------------------------------
9. W3-1: classification; threshold classifier output; logistic regression; (2019-3-24)

Reading: However, this method doesn't work well because classification is not actually a linear function;

1. W3-2: hypothesis Representation; sigmoid function; interpretation of hypothesis output; (2019-3-25)

Reading: hθ​(x)=0.7 gives us a probability of 70% that our output is 1; (2019-3-25)

1. W3-3: decision boundary; logic regression predict; non-linear decision boundary; (2019-3-25)

Reading: The decision boundary is the line that separates the area where y = 0 and where y = 1. It is created by our hypothesis function; (2019-3-25)

1. W3-4: Cost Function; convex function; logistic regression cost function; (2019-3-25)

Reading: If our correct answer 'y' is 0, then the cost function will be 0 if our hypothesis function also outputs 0. If our hypothesis approaches 1, then the cost function will approach infinity; (2019-3-25)

1. W3-5: Simplified Cost Function and Gradient Descent; fit parameters theta; (2019-3-25)

Reading: Notice that this algorithm is identical to the one we used in linear regression. We still have to simultaneously update all values in theta; (2019-3-25)

1. W3-6: Advanced Optimization; optimization algorithm; (2019-3-25)

Reading: We suggest that you should not write these more sophisticated algorithms yourself; (2019-3-25)

1. W3-7: Multiclass Classification: One-vs-all; one vs rest; (2019-3-25)

Reading: We are basically choosing one class and then lumping all the others into a single second class; (2019-3-25)

1. W3-8: The Problem of Overfitting; example, linear regression; under fit; reduce number of features; regularization; (2019-3-25)
2. W3-9: cost function; (2019-3-25)

Reading: If we have overfitting from our hypothesis function, we can reduce the weight that some of the terms in our function carry by increasing their cost; (2019-3-25)

1. W3-10: Regularized Linear Regression; gradient descent; non-invertibility; (2019-3-25)
2. W3-11: Regularized Logistic Regression; advanced optimization; (2019-3-26)

Reading: We can regularize logistic regression in a similar way that we regularize linear regression. As a result, we can avoid overfitting; (2019-3-26)

1. ---------------------------------------Chapter 4----------------------------------------
2. W4-1: Non-linear Hypotheses; non-linear classification; computer vision, car detection; (2019-3-27)
3. W4-2: Neurons and the Brain; neural networks; one learning algorithm hypothesis; sensor representation in the brain; (2019-3-27)
4. W4-3: Model Representation I; neuron in the brain; logistic unit; neural network; input layer, output layer; (2019-3-27)
5. W4-4: Model Representation II; neural network learning its own features; other network architecture; (2019-3-27)
6. W4-6: Examples and Intuitions I; non-linear classification example, XOR/XNOR; simple example, AND; example, OR; (2019-3-27)
7. W4-7: Examples and Intuitions II; negation; handwritten digit classification; (2019-3-28)

Reading: We can combine these to get the XNOR logical operator; (2019-3-28)

1. W4-8: Multiclass Classification; multiple output unit, one-vs-all; (2019-3-28)

Reading: To classify data into multiple classes, we let our hypothesis function return a vector of values; (2019-3-28)

1. ---------------------------------------Chapter 5----------------------------------------
2. W5-1: Cost Function; binary classification, multi-class classification; cost function; (2019-3-29)

Reading: We have added a few nested summations to account for our multiple output nodes; (2019-3-29)

1. W5-2: Backpropagation Algorithm; gradient computation; forward propagation; backpropagation algorithm; (2019-3-29)
2. W5-3: Backpropagation Intuition; forward propagation; what is backpropagation doing; (2019-3-29)
3. W5-4: Implementation Note, Unrolling Parameters; advanced optimization; (2019-3-30)

Reading: In order to use optimizing functions such as "fminunc()", we will want to "unroll" all the elements and put them into one long vector; (2019-3-30)

1. W5-5: Gradient Checking; numerical estimation of gradient; parameter vector theta; implementation note; (2019-3-30)
2. W5-6: Random Initialization; initial value of theta; zero initialization; random initialization, symmetry breaking; (2019-3-30)

Reading: Instead we can randomly initialize our weights for our Θ matrices using the following method; (2019-3-30)

1. W5-7: Putting It Together; training a neural network; (2019-3-30)
2. W5-8: Autonomous Driving; (2019-3-30)
3. ---------------------------------------Chapter 6----------------------------------------
4. W6-1: what to try next; debugging a learning algorithm; machine learning diagnostic; (2019-4-9)
5. W6-2: evaluating a hypothesis; data set; training/testing procedure for linear regression; training/testing procedure for logistic regression; (2019-4-9)
6. W6-3: Model Selection and Train/Validation/Test Sets; over-fitting example; model selection; evaluating your hypothesis; train/validation/test error; model selection; (2019-4-9)
7. W6-4: Diagnosing Bias vs. Variance; bias/variance; diagnosing bias (under fit) vs. Variance (over fit); (2019-4-9)
8. W6-5: Regularization and Bias/Variance; linear regression with regularization; choosing lambda; bias/variance as a function of lambda; (2019-4-9)
9. W6-6: Learning Curves; high bias; high variance; (2019-4-9)
10. W6-7: Deciding What to Do Next Revisited; debugging a learning algorithm; neural networks and over-fitting; (2019-4-12)
11. W6-8: Prioritizing What to Work On; building a spam classifier; (2019-4-12)
12. W6-9: Error Analysis; recommended approach; error analysis; the importance of numerical evaluation; (2019-4-12)
13. W6-10: Error Metrics for Skewed Classes; cancer classification example; precision/recall; (2019-4-13)
14. W6-11: Trading Off Precision and Recall; trading off precision and recall; F score; (2019-4-13)
15. W6-12: Data For Machine Learning; designing a high accuracy learning system; large data rationale; (2019-4-13)
16. ---------------------------------------Chapter 7----------------------------------------
17. W7-1: Optimization Objective; alternative view of logistic regression; SVM hypothesis; (2019-4-13)
18. W7-2: Large Margin Intuition; support vector machine; SVM decision boundary; SVM decision boundary, linearly separable case; large margin classifier in presence of outliers; (2019-4-13)
19. W7-3: Mathematics Behind Large Margin Classification; vector inner product; SVM decision boundary; (2019-4-13)
20. W7-4: Kernels I; non-linear decision boundary; kernel; kernels and similarity; example; (2019-4-13)
21. W7-5: Kernels II; choosing the landmarks; SVM with kernels; SVM parameters; (2019-4-13)
22. W7-6: Using An SVM; use SVM software package; kernel (similarity) functions; other choices of kernel; multi-class classification; logistic regression vs SVMs; (2019-4-18)
23. ---------------------------------------Chapter 8----------------------------------------
24. w8-1: Unsupervised Learning: Introduction; supervised learning; unsupervised learning; application of clustering; (2019-4-18)
25. W8-2: K-Means Algorithm; k-means for non-separated clusters; (2019-4-18)
26. W8-3: Optimization Objective; k-means optimization objective; K-means algorithm; (2019-4-18)
27. W8-4: Random Initialization; k-means algorithm; random initialization; local optima; (2019-4-18)
28. W8-5: Choosing the Number of Clusters; what is the right value of K; elbow method; choosing the value of K; (2019-4-18)
29. w8-6: Motivation I: Data Compression; data compression; (2019-4-18)
30. W8-7: Motivation II: Visualization; data visualization; (2019-4-18)
31. W8-8: Principal Component Analysis Problem Formulation; PCA problem formulation; PCA is not linear regression; (2019-4-19)
32. W8-9: Principal Component Analysis Algorithm; data preprocessing; PCA algorithm; PCA algorithm summary; (2019-4-19)
33. W8-10: Reconstruction from Compressed Representation; reconstruction from compressed representation; (2019-4-19)
34. W8-11: Choosing the Number of Principal Components; choosing k; (2019-4-19)
35. W8-12: Advice for Applying PCA; supervised learning speedup; application of PCA; bad use of PCA, to prevent overfitting; PCA is sometimes used where it shouldn’t be; (2019-4-19)
36. ---------------------------------------Chapter 9----------------------------------------
37. W9-1: Problem Motivation; anomaly detection example; density estimation; anomaly detection example; (2019-4-20)
38. W9-2: Gaussian Distribution; Gaussian Distribution; Gaussian distribution example; parameter estimation; (2019-4-20)
39. W9-3: Algorithm; density estimation; anomaly detection algorithm; anomaly detection example; (2019-4-20)
40. W9-4: Developing and Evaluating an Anomaly Detection System; the importance of real-number evaluation; aircraft engine motivating example; algorithm evaluation; (2019-4-20)
41. W9-5: Anomaly Detection vs. Supervised Learning; anomaly detection vs supervised learning; (2019-4-20)
42. W9-6: Choosing What Features to Use; non-gaussian features; error analysis for anomaly detection; monitoring computers in a data center; (2019-4-20)
43. W9-7: Multivariate Gaussian Distribution; motivating example, monitoring machines in a data center; Multivariate Gaussian Distribution; Multivariate Gaussian examples; (2019-4-20)
44. W9-8: Anomaly Detection using the Multivariate Gaussian Distribution; anomaly detection with the multivariate Gaussian; relationship to original model; original model vs multivariate Gaussian; (2019-4-21)
45. W9-9: Problem Formulation; example, predicting movie ratings; (2019-4-21)
46. W9-10: Content Based Recommendations; content-based recommender systems; problem formulation; optimization objective; optimization algorithm; (2019-4-21)
47. W9-11: Collaborative Filtering; problem motivation; optimization algorithm; collaborative filtering; (2019-4-21)
48. W9-12: Collaborative Filtering Algorithm; collaborative filtering optimization objective; collaborative filtering algorithm; (2019-4-21)
49. W9-13: Vectorization: Low Rank Matrix Factorization; collaborative filtering; finding related movies; (2019-4-21)
50. W9-14: Implementational Detail: Mean Normalization; users who have not rated any movies; mean normalization; (2019-4-22)
51. ---------------------------------------Chapter 10----------------------------------------
52. W10-1: Learning With Large Datasets; machine learning and data; learning with large datasets; (2019-4-23)
53. W10-2: Stochastic Gradient Descent; linear regression with gradient descent; batch gradient descent, stochastic gradient descent; stochastic gradient descent; (2019-4-23)
54. W10-3: Mini-Batch Gradient Descent; mini-batch gradient descent; (2019-4-23)
55. W10-4: Stochastic Gradient Descent Convergence; checking for convergence; stochastic gradient descent; (2019-4-23)
56. W10-5: Online Learning; online learning; other online learning example; (2019-4-23)
57. W10-6: Map Reduce and Data Parallelism; map reduce; map reduce and summation over the training set; multi-core machines; (2019-4-23)
58. ---------------------------------------Chapter 11----------------------------------------
59. w11-1: Problem Description and Pipeline; the photo OCR problem; photo OCR pipeline; (2019-4-23)
60. W11-2: Sliding Windows; text detection, pedestrian detection; supervised learning for pedestrian detection; sliding window detection; text detection; 1D sliding window for character segmentation; photo OCR pipeline; (2019-4-23)
61. W11-3: Getting Lots of Data and Artificial Data; character recognition; artificial data synthesis for photo OCR; synthesizing data by introducing distortions; discussion on getting more data; (2019-4-23)
62. W11-4: Ceiling Analysis: What Part of the Pipeline to Work on Next; estimating the errors due to each component; another ceiling analysis example; (2019-4-23)
63. W11-5: Summary and Thank You; main topics; (2019-4-23)
64. -