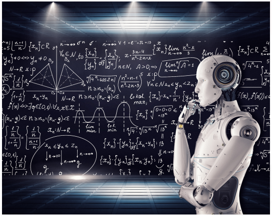
Introduction, Terminology, and Review

Dana Golden, Lilia Maliar



Data Science and Machine Learning - **November 30, 2024**

Presentation Outline

Who am I? Dana Golden

- Third year PhD Student in economics at SBU
- MS degrees in economics and analytics with a concentration in machine learning
- Worked as an economist for USDA, FERC, and data scientist for Treasury
- Research focuses on electricity markets and energy transition
- Ask me about: Natural language processing, computer vision, deep learning, reinforcement learning

Office Hours

- Office hours on Zoom and in person at my office (S630)
- 3:30-5:30 PM Monday, 8-10 AM Friday

Overview of Course

- Course covers broad introduction to machine learning including:
  - Optimization, gradient descent, Newton's method
  - Linear regression from a machine learning perspective
  - Dimensionality reduction: SVD, PCA
  - Overfitting, underfitting, cross-validation
  - Supervised learning: logistic regression, neural networks, SVM, decision trees, random forest
  - Unsupervised learning: K-means, K-mediods
  - Large-scale machine learning
  - Reinforcement learning
- Grades based on participation (5%), problem sets (15%), 2 midterm exams (20% each), and a final (40%)
- Problem sets will be in Python
- Tests will be either proctored online or proctored in person

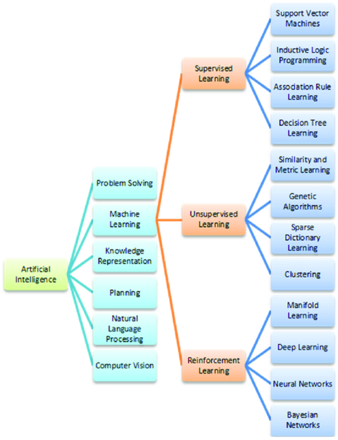
Plan for Recitations

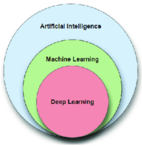
- In person and online: 5-7 PM Wednesdays and 8-10 AM Thursdays
- Math review and extension
- Coding demonstration
  - All recitations will come with sample code
- Work practice problems
- Answer questions from students
  - If you have a question you want reviewed, just send beforehand

Plan for Recitations

- In person and online: 5-7 PM Wednesdays and 8-10 AM Thursdays
- Math review and extension
- Coding demonstration
  - All recitations will come with sample code
- Work practice problems
- Answer questions from students
  - If you have a question you want reviewed, just send beforehand
- Anything else you want included?

Overview of Machine Learning





How Machine Learning is Used in Economics

- Creating quick classifiers of unstructured data
- Working with text or image data
- Prediction tasks or sorting through variables
- Solving value-iteration problems using deep learning

How Machine Learning is Used in Economics

- Creating quick classifiers of unstructured data
- Working with text or image data
- Prediction tasks or sorting through variables
- Solving value-iteration problems using deep learning
- Potential for explosive growth over time

The Economist's Comparative Advantage

- Economists are not the best programmers. So why hire them to do machine learning?

The Economist's Comparative Advantage

- Economists are not the best programmers. So why hire them to do machine learning?
- Economists know data

The Economist's Comparative Advantage

- Economists are not the best programmers. So why hire them to do machine learning?
- Economists know data
- Economists know the theories that allow them to make sense of data

The Economist's Comparative Advantage

- Economists are not the best programmers. So why hire them to do machine learning?
- Economists know data
- Economists know the theories that allow them to make sense of data
- Economists know causality

The Economist's Comparative Advantage

- Economists are not the best programmers. So why hire them to do machine learning?
- Economists know data
- Economists know the theories that allow them to make sense of data
- Economists know causality
- Economists have a rigorous approach to statistics missing from machine learning

Machine Learning vs. Econometrics

- In a lot of ways, machine learning is just a rebranding of econometrics and classical statistics but there are some differences

Machine Learning vs. Econometrics

- In a lot of ways, machine learning is just a rebranding of econometrics and classical statistics but there are some differences
- Machine learning focuses on prediction and implementability, economists care about causality and statistical rigor

Machine Learning vs. Econometrics

- In a lot of ways, machine learning is just a rebranding of econometrics and classical statistics but there are some differences
- Machine learning focuses on prediction and implementability, economists care about causality and statistical rigor
- Many economic models have analytical solutions, machine learning models rarely do

Common Terminology

- Loss/cost function: The objective function to be optimized during training
- Train-test split: Before training a model, we hold out some data for testing. Economists don't do this. Why do machine learning people?
- Overfitting: Overfitting is when a model is too complex and fits training data better than testing data
- Bias-variance tradeoff: Bias is how wrong an estimator is on average, variance is how volatile an estimator is. The right estimator contains the optimal amount of bias and variance.
- Cross-validation: Cross-validation is the process of finding parameters for the model using a subset of training data

Questions?

Partial Derivatives

$$f(x, y) = x^2 y + y^3 \tag{1}$$

Gradient

- Gradient represents the vector derivative of a function
- Important for optimization

$$g(x, y, z) = xe^{yz} \tag{2}$$

Lagrangian Optimization

- Lagrangian optimization is a technique for optimizing a multivariate function subject to constraints
- Given the following function and constraints, find critical points

$$h(x, y) = 3x^2 + 2xy + y^2 s.t. x + y = 1 \tag{3}$$

Chain Rule

- Chain rule is at the bedrock of optimization

$$f(u,v) = u^2 + v^2 \tag{4}$$

$$u = x + y \tag{5}$$

$$v = x - y \tag{6}$$

Hessian Matrix

- The Hessian Matrix is the matrix of second derivatives
- Very important when maximizing function or finding

$$f(x, y) = 4x^3 + 2xy + y^2 \tag{7}$$

Basic Probability Concepts

- Expected Value: $\sum_{i=1}^{n} x_i P(x_i)$
- Variance: $E(x^2) - E(x)^2 = \frac{\sum (x_i - \mu)^2}{n}$, . Will this always be positive?
- Correlation: $\rho = \frac{\sigma_{xy}}{\sqrt{\sigma_x^2 \sigma_y^2}} = \frac{\sum (x_i - \mu_x)(y_i - \mu_y)}{\sqrt{\sum (x_i - \mu_x)^2 \sum (y_i - \mu_y)^2}}$
- Independence: $E(XY) = E(X)E(Y)$

Basic Probability Problems

- Let X_1 and X_2 be independent continuous random variables uniformly distributed from 0 to 1. Let $x = max(X_1, 2X_2)$
- Find $E(X)$
- Find $Var(X)$
- Find $Cov(X_1, X_2)$

Baysesian vs. Frequentist

- Frequentist interpretations base analysis on the data and use the data as empirical observations of a fixed distribution
- Bayesian interpretations utilize a prior belief and see data as a way to update beliefs
- Most machine learning is frequentist

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \tag{8}$$

$$\frac{P(A \cap B)}{P(A)} \tag{9}$$

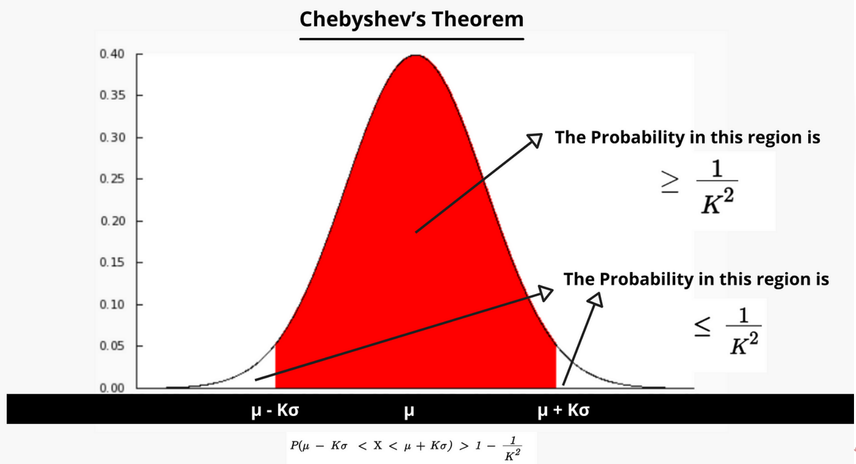
Bayesian Sample Problem

A crime has been committed. Assume that the crime was committed by exactly one person, that there are 1000 people who could have committed the crime and that, in the absence of any evidence, these people are all equally likely to be guilty of the crime. A piece of evidence is found. It is judged that this evidence would have a probability of 0.99 of being observed if the crime were committed by a particular individual, A, but a probability of only 0.0001 of being observed if the crime were committed by any other individual.

Find the probability, given the evidence, that A committed the crime.

Chebyshev's Theorem

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2} \tag{10}$$



Law of Large Numbers

Then, as $n \to \infty$, the sample mean $\langle x \rangle$ equals the population mean μ of each variable.

$$\begin{aligned}
\langle X \rangle &= \left\langle \frac{X_1 + \ldots + X_n}{n} \right\rangle \\
&= \frac{1}{n}\left(\langle X_1 \rangle + \ldots + \langle X_n \rangle\right) \\
&= \frac{n\,\mu}{n} \\
&= \mu.
\end{aligned}$$

In addition,

$$\begin{aligned}
\operatorname{var}(X) &= \operatorname{var}\left(\frac{X_1 + \ldots + X_n}{n}\right) \\
&= \operatorname{var}\left(\frac{X_1}{n}\right) + \ldots + \operatorname{var}\left(\frac{X_n}{n}\right) \\
&= \frac{\sigma^2}{n^2} + \ldots + \frac{\sigma^2}{n^2} \\
&= \frac{\sigma^2}{n}.
\end{aligned}$$

Central Limit Theorem

5.8 Theorem (The Central Limit Theorem (CLT)). *Let* $X_1, \ldots, X_n$ *be* IID *with mean* μ *and variance* σ^2. *Let* $\overline{X}_n = n^{-1}\sum_{i=1}^{n} X_i$. *Then*

$$Z_n \equiv \frac{\overline{X}_n - \mu}{\sqrt{\mathbb{V}(\overline{X}_n)}} = \frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \rightsquigarrow Z$$

where $Z \sim N(0,1)$. *In other words,*

$$\lim_{n\to\infty} \mathbb{P}(Z_n \leq z) = \Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx.$$

Maximum Likelihood Estimator

- MLEs maximize the probability of an estimator
- IID distributions are multiplied, the logarithm is taken, and the maximum is found
- Makes most assumptions about distributions
- Consistent and efficient but not certain to be unbiased

OLS via ML

$$y_t = X_t\beta + u_t, u_t \sim iidN(0, \sigma^2) \tag{11}$$

$$L(y|\beta, \sigma) = \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi\sigma}} exp\{\frac{-1}{2\sigma}(y - X_t\beta)^2\} \tag{12}$$

$$ln(L(y|\beta, \sigma)) = \sum_{t=1}^{T} \frac{-1}{2} ln(2\pi) - ln(\sigma) - \frac{1}{2\sigma}(y_t - X_t\beta)^t \tag{13}$$

- This is maximized by minimizing the sum of squared errors

Method of Moments Estimator

- Method of moments estimators use the empirical moments as the true moments and solve for estimators that minimize the difference
- Method of moments estimators (MOMs) make fewer assumptions
- Unbiased but generally less efficient

Unbiased and Consistent Estimators

- Unbiased estimators have an expectation that is asymptotitcally equal to the true value
- Consistent estimators converge probabilistically to the true value
- Unbiasedness is a small-sample property, consistency is strictly a large-sample property

Thank You So Much!

List of References