Machine Learning

Prof. Sergei Gleyzer

Week 2

PH451 PH551

January 14, 2025

Announcements

- HW #1 Reading Assignment out today
  - Due next Tuesday 01/21 at 1pm
- Github repository
  - Hands-on exercises
- Please join the Slack channel
  - Questions, technical assistance

Outline

- **Relevant concepts**
  - **Probability and Statistics**
  - **Mathematics and Optimization**

Relevant Concepts

Optimization

Optimization is a branch of mathematics aiming to solve the following problem:

- **How to find elements that maximize or minimize a given function**
- **Many problems can be cast in terms of mathematical optimization**
  - **including various machine learning tasks**
  - **i.e. minimize misclassification of instances**

Optimization

Given real-valued function f: $R^p \rightarrow R$

$$\underset{\boldsymbol{x} \in \mathbb{R}^p}{\text{minimize}}\ f(\boldsymbol{x})$$

f is the objective function

- also loss function or cost function

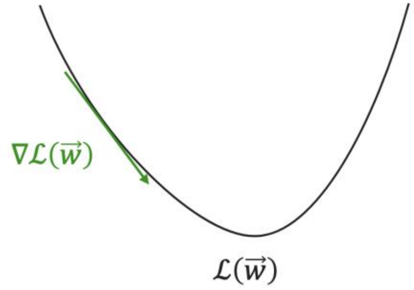
Gradient

Gradient: vector whose components are partial derivatives of a function

if **f** is differentiable: $\nabla f(w)$

Gradient

$$\nabla \mathcal{L}(\vec{w}) = \left(\frac{\partial \mathcal{L}(\vec{w})}{\partial w_1}, \ldots, \frac{\partial \mathcal{L}(\vec{w})}{\partial w_d} \right)$$



A. Cauchy

Key questions

Dimensionality: $R^p \rightarrow R$ (p, number of features)

Is f convex or not?

Is f smooth?

Useful

Chain rule in derivatives:

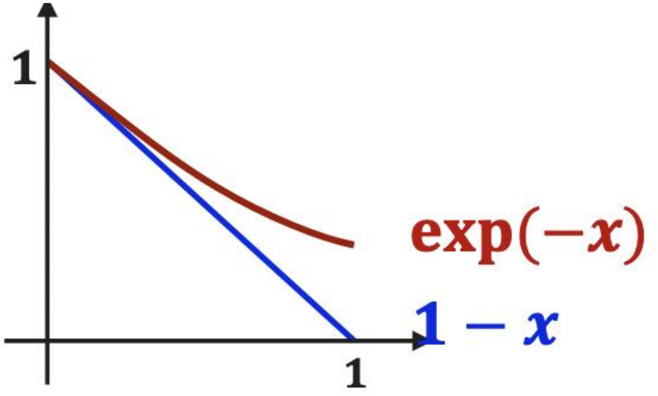
$$\frac{\partial f}{\partial x} = \frac{\partial f}{\partial g(x)} \cdot \frac{\partial g(x)}{\partial h(x)} \cdot \frac{\partial h(x)}{\partial x}$$

Example

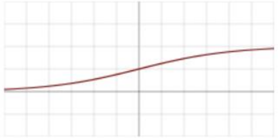
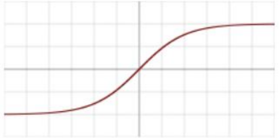
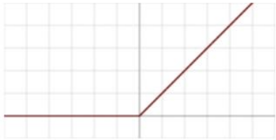
Chain rule in derivatives:

$$\frac{\partial f}{\partial x} = \frac{\partial f}{\partial g(x)} \cdot \frac{\partial g(x)}{\partial h(x)} \cdot \frac{\partial h(x)}{\partial x}$$

Lemma: $(1 - \epsilon) \leq e^{-\epsilon}$



Useful Functions

Name	Function	Gradient	Graph
Binary step	$\mathrm{sign}(x)$	$\begin{cases} 0 & x \neq 0 \\ N/A & x = 0 \end{cases}$	
sigmoid	$\sigma(x) = \dfrac{1}{1 + \exp(-x)}$	$\sigma(x)(1 - \sigma(x))$	
Tanh	$tanh(x) = \dfrac{\exp(x) - \exp(-x)}{\exp(x) + \exp(-x)}$	$\left(1 - tanh(x)\right)^2$	
Rectified Linear (ReLu)	$relu(x) = \max(x, 0)$	$\begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases}$	

Supervised Learning

R^d: d-dimensional input (feature or variable) space

- $\mathbf{x}_i$ **:** input vector, i^{th} sample (feature vector)
- Y_i **:** label, i^{th} sample

Goal: Find a function h ($\mathbf{x}_i$) that approximates Y_i

Supervised Learning

Examples:

- Classify if a patient has a condition (for ex. cancer) based on measured characteristics
  - features of nuclei present in images
  - One of the classic ML datasets

- Classify if an email you received is spam (spam detection) based on keywords and their frequency
  - Free, money, offer, business, send….

Hypothesis Testing

Important for making decisions based on experimental data:

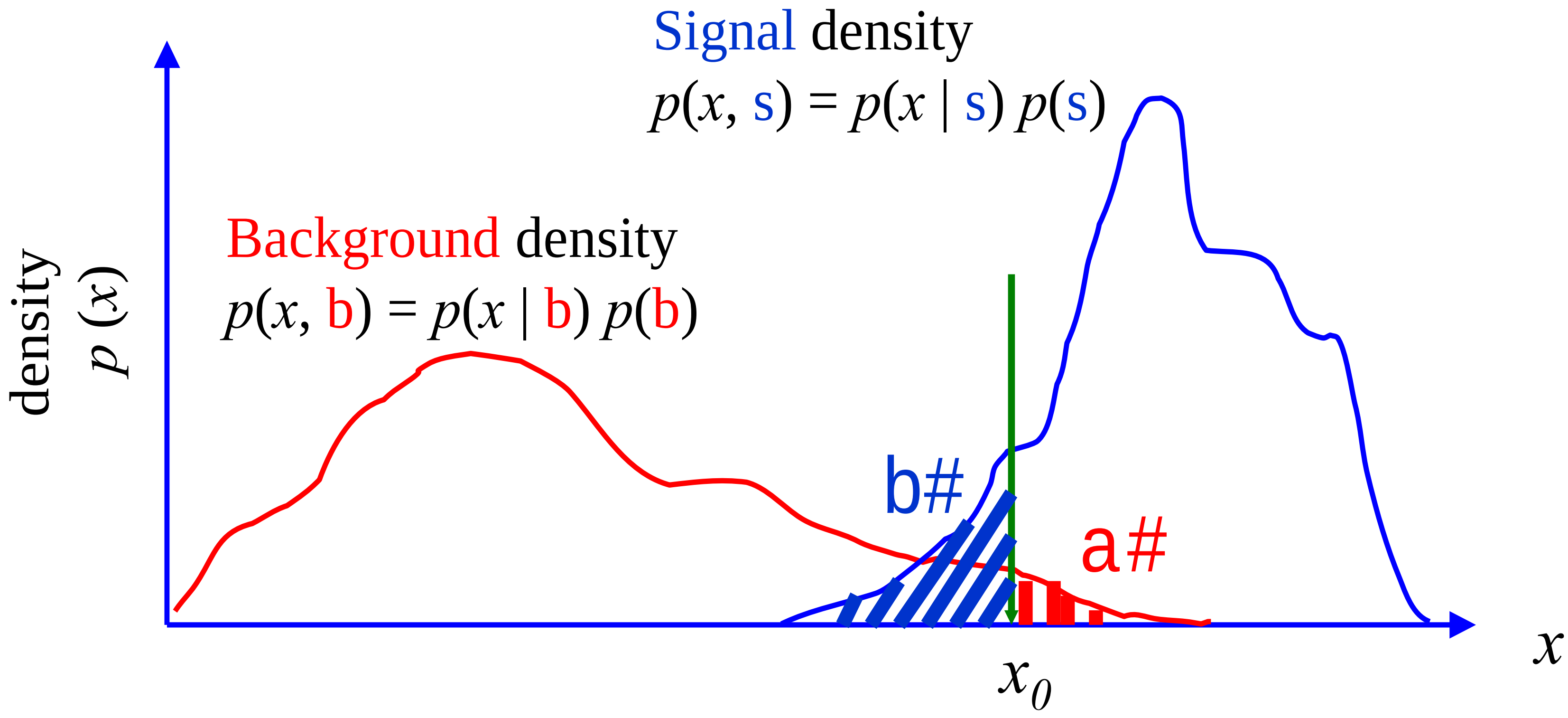
- Null hypothesis (H0) – basic assumption
- Alternative hypothesis (H1) – contradicts the assumption

Hypothesis test: keep or reject the null hypothesis

Set a threshold on level of significance

- i.e. p-value < threshold reject the null hypothesis

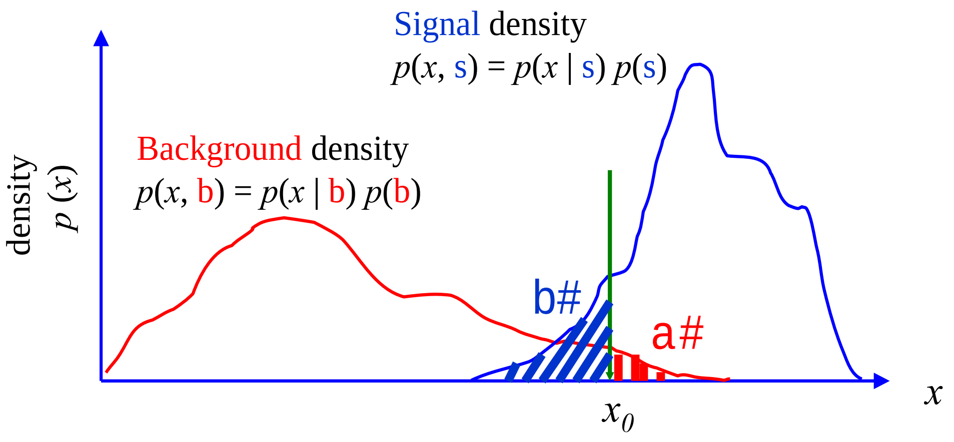
Classification



Hypothesis Testing

Type I error (α): reject the null hypothesis when it is true

Type II error (β): accept the null hypothesis when it is false



Loss Function

Given a loss function, we can try to find the $f(x_i)$ that minimizes it

- Much of machine learning deals with how to achieve this **minimization** most efficiently

- Key: how well does the solution generalize to instances the model has not seen during training
  - We will come back to this in later lectures

Maximum Likelihood Estimate

For estimating parameter values of a statistical model

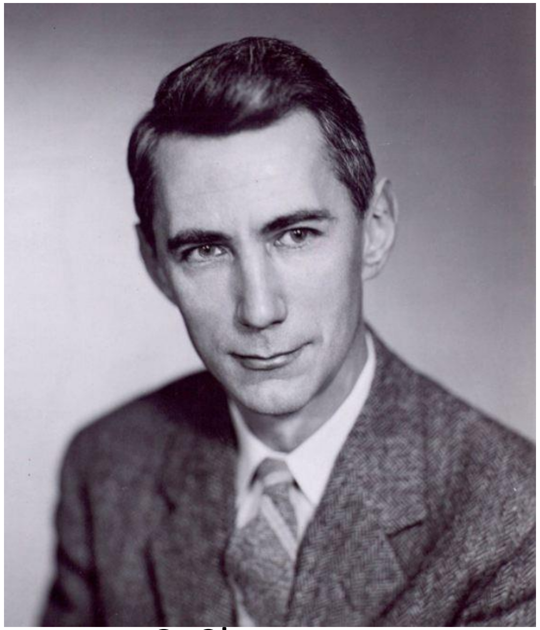
- Find those parameter values that maximize the likelihood of making the observations

Entropy

Information theory concept:

- Measure of “impurity” or “uncertainty” of data
- Completely homogeneous sample: entropy = 0

$$Entropy\,(p) = -\sum_{i=1}^{N} p_i \log_2 p_i$$

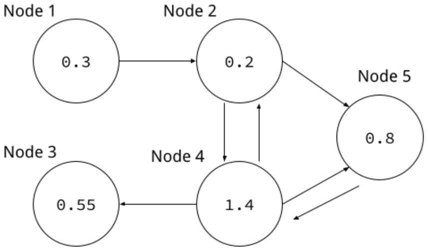


C. Shannon

Class Introduction (cont.)

Colin Crovella - Introduction

- 5th year Graduate student at UA
- Undergraduate
  - BS in Physics at University of Virginia
  - Minor in Computer Science
- Research Interests
  - Collider Physics
  - Graph Neural Networks
  - Neural Networks on FPGAs
- Other Interests
  - Science Communication
  - Video games
  - World cuisine



Today’s plan

- **Python Primer**
- **Start Hands-on exercise #0**
  - **Individual**
  - **Python basics**
  - **Work on it this week**
  - **Due on Tue 01/21 at 1pm**