Statistical Techniques for Data Science (& Robotics)

Week 1

Objectives (for today)

(1) to learn about the course, including

- How to pass / fail
- How to get an “A”
- connections to 5+ courses from your program

(2) to recall Probability theory

(3) to do short recall basics of Statistics

Team and Communication

(1) PI: Vladimir Ivanov

- @nomemm
- room 475

(2) Teaching Assistants:

- Zamira Kholmatova
- Vladimir Bazilevich
- Zlata Shchedrikova

Syllabus

briefly

(1) “Classical” Statistics: Tests, Hypotheses

(2) Statistical view at Machine Learning models

(3) Non-parametric Statistics

Midterm

(4) Bayesian Statistics

(5) Sampling, MCMC

(6) Bandit algorithms

Structure

(1) **12-14** Lectures (2 hours/week) + Quizzes (10 min. on a lecture)

(2) **3** Assignments (aka Homework, up to 4 hours / week)

(3) **12-14** Labs (2 hours/week)

(4) **1** Case Study instead of the Midterm

(5) **1** Final Exam

(6) **0...n** retakes

Books

(1) Bruce, Peter,Bruce, Andrew. Practical Statistics for Data Scientists: 50 Essential Concepts. O'Reilly Media. — Simple and short, an **Intro-level book**

(2) Bishop Christopher. Pattern Recognition and Machine Learning. Springer, 2006. — 738 p. — **CoreBook, hard, but worth**

(3) Introduction to Mathematical Statistics. By ROBERT V. HOGG AND ALLEN. B. CRAIG 4th Edition — **MathStats book**

(4) Probability & statistics for engineers & scientists/Ronald E. Walpole . . . [et al.] — 9th ed. — **your ProbStat book (from the Fall semester)**

Grading

(1) Assignments: **30** % (aka Homeworks)

(2) Midterm : **30** % (Case study)

(3) Final Exam: **30** % (Written + Oral)

(4) Labs (grading during labs): **10 pts** (0, 1/2, 1 scale; max 1 point per lab)

86 - 100	A
70 - 85	B
55 - 69	C
0 - 54	D

Tools

(1) Pen and Paper

(2) R / Python (ver. 3+)

Course Prerequisites

(1) Linear Algebra, Mathematical analysis courses

(2) A course on Probability Theory

How to success?

(1) Assignments and Labs:

- work hard (individually) + office hours
- visit Labs to have enough practice with tools
- visit Labs to solve pen and paper problems

(3) Exam:

- read the books + office hours
- solve exercises from books at home

Discussion

(1) Which courses this course is connected to?

Break, 5 min.

Example

Fitting of a curve

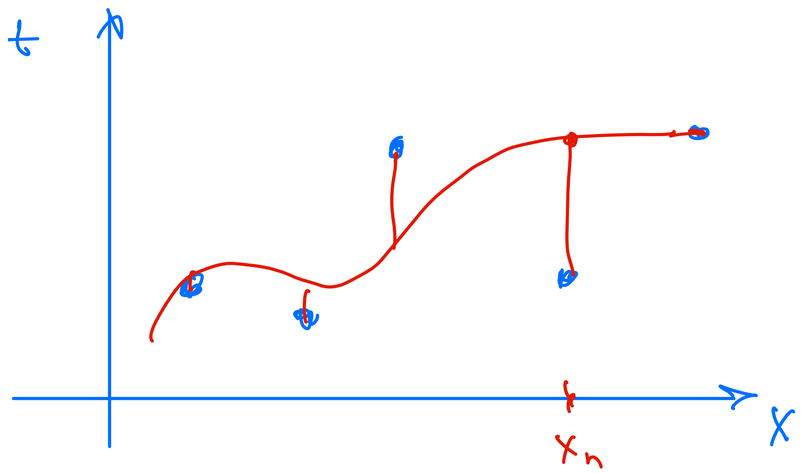
input variable x to predict t (target)

We generate data points

$$\sin(2\pi x) + \text{noise} = t(x)$$

train set N observations

$$X = (x_1, \dots x_N)^T \qquad t = (t_1, \dots, t_N)^T$$



predict $\hat{t}$ using $\hat{x}$

find a model for the trend (sin)

and deal with the noise

Normal Distrib-n.

Due to the noise we have uncertainty when predict $\hat{t}$ given $\hat{x}$

+ Prob stat helps to measure U.
+ Decision Theory helps to make optimal decisions (given some criteria)

1) model is a polynomial

if 5 points $\Rightarrow M=4$

$$y(x, w) = \sum_{j=0}^{M} w_j x^j \qquad j = 0, \dots, M$$

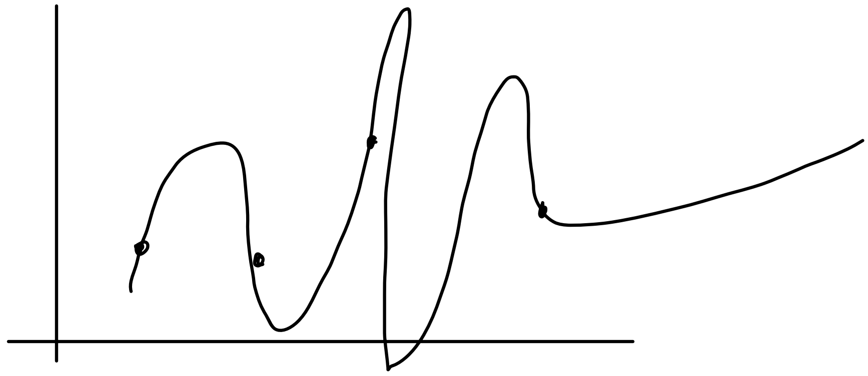
$y(x, w)$ – is linear w.r.t. $\underline{w_j}$

Errors:

$$E(w) = \frac{1}{2} \sum_{n=1}^{N} \left(y_n(x_n, w) - t_n \right)^2$$

$\min_{w}$

We can select w, but also M



$$\tilde{E}(w) = \frac{1}{2}\sum_{n=1}^{N}\left(y(x_n, w) - t_n\right)^2 + \underbrace{\frac{\lambda}{2}\|w\|^2}_{\text{Reg-n.}}$$

Recap of Probability Theory

Major Concepts

What is …?

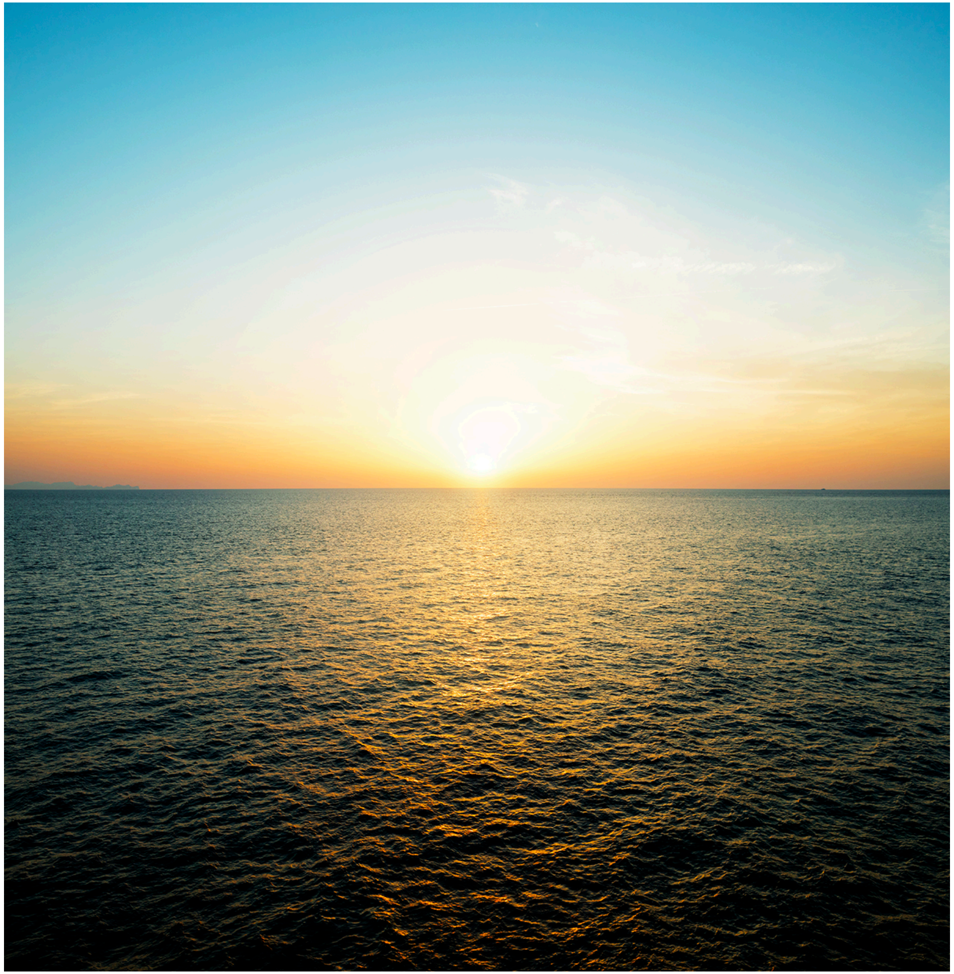
(1) Random Variable

(2) Expected value, variance

(3) Probability Density Function

(4) CDF

(5) Bayes Theorem



Random variables (r. v.)

(1) What are examples of random variables (YES/NO)?

N A. Winning a lottery

B. Choosing a green ball from an urn with a large mixture of red and black balls

Y C. Total value from a roll of two dice

N D. Two people in a classroom sharing the same birthday

Y E. The average exam score of a class if every student guesses answers

Y F. Winings from a game with a $1 gain/loss for each head/tail coin flip in a series of 10 flips

Random variables

(1) What are examples of random variables?

A. Winning a lottery

B. Choosing a green ball from an urn with a large mixture of red and black balls

C. **Total value from a roll of two dice**

D. Two people in a classroom sharing the same birthday

E. **The average exam score of a class if every student guesses answers**

F. **Number of winings from a game with a $1 gain/loss for each head/tail coin flip in a series of 10 flips**

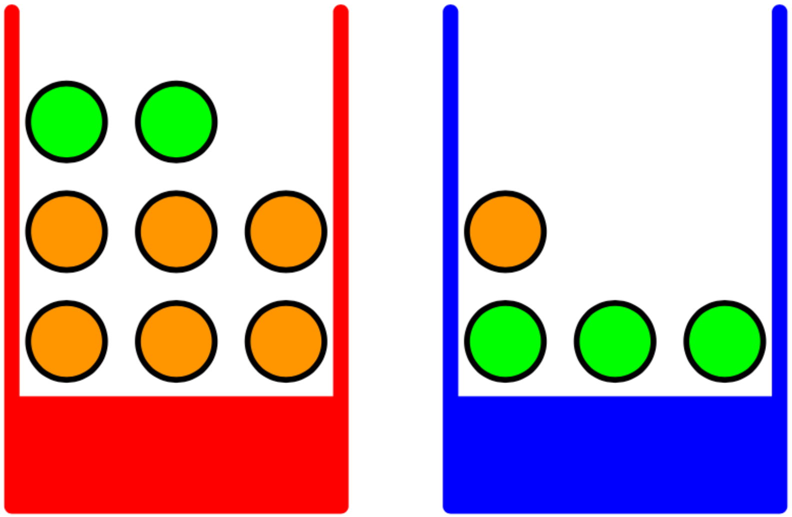
2 Rules of Probability

Apples and Oranges

(1) A - fruit type

(2) B - box color

- Q: probability of …
- given that we have chosen an orange, what is the probability that the box we chose was the blue one?

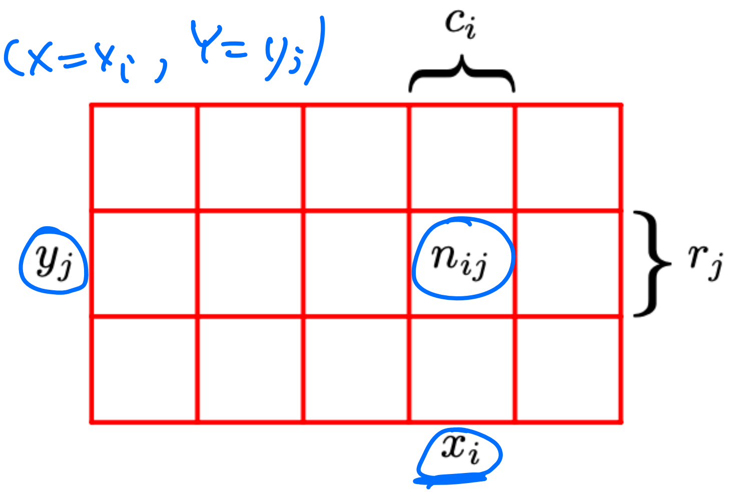


General Case

Sum rule

X, Y - r.v.

$\frac{n_{ij}}{N} = p(X=x_i, Y=y_j)$



(1) N is a number of all experiments

(2) n_{ij} is a number of trials when

- $X = x_i$ and $Y = y_j$

(3) c_i is the sum of i-th column

(4) $P(X=x_i) = \dots$ $\sum_j p(x,y) = \sum_j p(X=x_i, Y=y_j)$

(5) Sum rule:

- $P(X = x_i) = \dots$

X can take $x_1, \dots, x_M$

Y can take $y_1, \dots, y_L$

General Case

product rule

(1) n_{ij} is a number of trials when

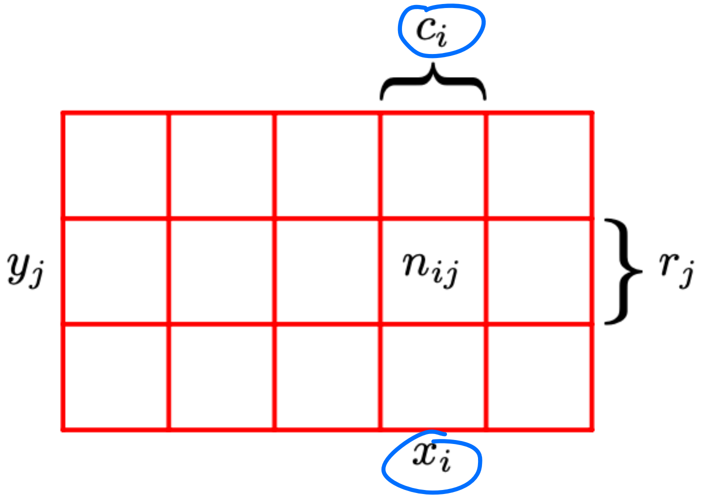
- $X = x_i$ and $Y = y_j$

(2) c_i is the sum of i-th column

(3) $P(X{=}x_i) = c_i/N$

(4) Product rule:

- $P(X = x_i, Y = y_j) = \boxed{\dfrac{n_{ij}}{c_i}} \cdot \boxed{\dfrac{c_i}{N}} = \ldots$ $P(Y = y_j \mid X = x_i) \cdot P(x_i)$



2 Rules of Probability

sum rule $$p(X) = \sum_{Y} p(X, Y)$$

product rule $$p(X, Y) = p(Y|X)p(X).$$

Probability Distributions (of r.v.)

(1) **A random variable** quantifies chance events, and

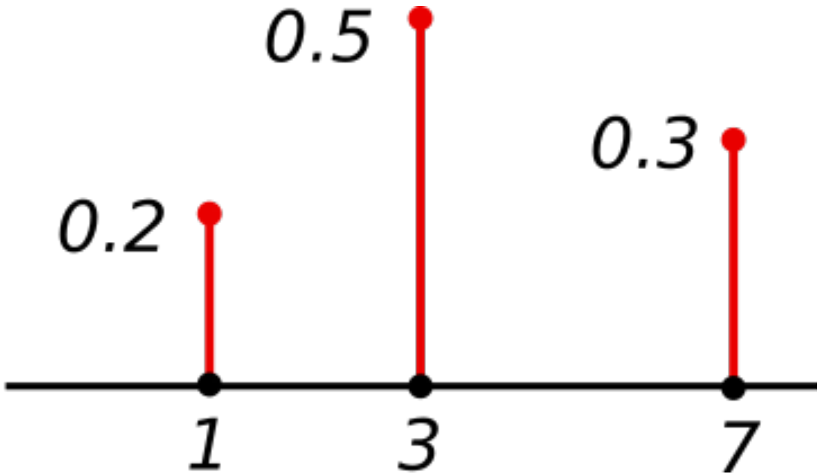
(2) **its probability distribution** assigns a likelihood to each of its (r.v.) values.

(3) Depending on nature of event the r.v. and distribution can be:

- discrete
- continuous

Probability mass function

For discrete random variable X: PMF



Probability density function

For continuous random variable X: PDF

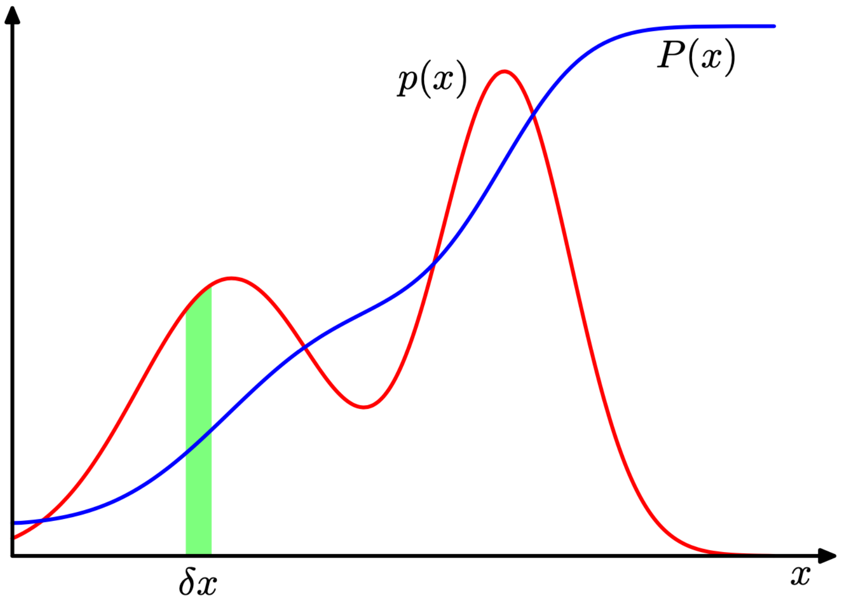
$f_X(x)$ is a probability density function for r.v. X

$$P[a < X \leq b] = \int_a^b f_X(x)\, dx$$

x is just a value (argument) from the domain of $f_X(x)$

Question

$$\begin{aligned} p(x) &\geqslant 0 \\ \int_{-\infty}^{\infty} p(x)\,\mathrm{d}x &= 1. \end{aligned}$$



What about the range of a pdf $p(x)$?

Cummulative distribution function

CDF

(1) if $f_X(x)$ is continuous at x:

$$F_X(x) = \int_{-\infty}^{x} f_X(u)\, du$$

$$f_X(x) = \frac{d}{dx} F_X(x)$$

$$F_X(x)' = f_X(x)$$

Expected value and Variance

(1) Given X is a random variable

(2) Well, try to understand the formulas:

$$E[X] = \sum_{n \in X's \quad range} nP(X = n)$$

$$V[X] = Var[X] = E[(X - E[X])^2]$$

Show (as an exercise) that

$$Var[X] = E[X^2] - (E[X])^2$$

Expected value (continuous r.v.)

(1) Well, try to understand the formula:

$$E[X] = \int_{x \in S} xp(x)dx$$

S is range of random variable X

Expected value and its Properties

- $E[X] = \sum xp(x)$

- $E[X] = \int xp(x)dx$

- $E[X + c] = E[X] + c$
- $E[cX] = cE[X]$
- $E[X + Y] = E[X] + E[Y]$

Conditional Expectation: discrete case

(1) The expected value of X given the event $Y = y$:

(2) $E[X \mid Y = y] = \sum_{x} xP(X = x \mid Y = y)$

HOMEWORK exercise:
What is the expected number of "heads" flips in 5 flips of a fair coin, given that the number of "heads" flips is greater than 2?

Conditional Expectation: cont. case

(1) The expected value of X with pdf $p(x)$

(2) Let S be the range of X given the event $Y = y$:

$$E[X \mid Y = y] = \frac{\int_{x \in S} xp(x)dx}{P(Y = y)}$$

HOMEWORK exercise: Let X be a continuous random variable

with density function $p(x) = 2x; 0 < x < 1$. What is $E[X \mid X < \frac{1}{2}]$?

Distributions

Bernoulli and Binomial

(1) Any random experiment whose outcome can be classified as either a success or a failure is called a **Bernoulli trial**.

(2) If you run T trials (and probability of success is **p**), then the number of successes is a r.v. N that has a discrete distribution.

(3) Binomial distribution

$$P(N = k) = \binom{T}{k} p^k (1 - p)^{T-k}$$

Poisson distribution

(1) Your phone glitches randomly

(2) with certain rate of k failures per second.

(3) X is the number of glitches in a time period *t* since release

(4) X is a r.v. that has Poisson distribution

$$P(X=n)=\frac{\lambda^n}{n!}e^{-\lambda}, n=1,2,3,\ldots$$

$$\lambda = kt$$

Poisson variables/distribution

(1) Find

$$P(X = n) = \frac{\lambda^n}{n!} e^{-\lambda}, n = 1,2,3,...$$

$$\lambda = kt$$

$$E[X] = ?$$

$$Var[X] = ?$$

Exponential

(1) For instance, the time t it takes for the decay of a radioactive carbon-14 atom into stable nitrogen-14 is distributed **exponentially**

$$f(t) = ae^{-at}, t \geq 0, a > 0$$

What's the probability that a carbon-14 atom lasts at least as long as $\frac{1}{a}$?

About Normal Distribution

Univariate case

$$\mathcal{N}\left(x|\mu,\sigma^2\right) = \frac{1}{(2\pi\sigma^2)^{1/2}}\exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$

$\beta = \dfrac{1}{\sigma^2}$ is a precision

Normal Distribution

Multivariate case (aka D-dimensional vector r.v.)

$$\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\boldsymbol{\Sigma}|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\}$$

Σ is a $D \times D$ covariance matrix

Bayesian Probabilities

(1) Philosophically, all probabilities are conditional probabilities.

Bayes' Theorem

(1) formula describes how to **update** the probabilities of hypotheses (H) when given evidence (E).

$$P(H \mid E) = \frac{P(E \mid H)}{P(E)} P(H)$$

$$f_y(y \mid X = x) = \frac{f_X(x \mid Y = y)}{f_X(x)} f_Y(y)$$

Bayes' Theorem

(1) prior distribution,

(2) posterior distribution, and

(3) likelihood ratio

$$P(H \mid E) = \frac{P(E \mid H)}{P(E)} P(H)$$

Alternative view

(1) a model, with parameters $\mathbf{w}$

(2) data observed: $\mathbf{D}$

(3) Likelihood of data, given a model, **likelihood function**: $p(\mathbf{D} \mid \mathbf{w})$

(4) **Prior** probability : $p(\mathbf{w})$

(5) **Posterior** probability $p(\mathbf{w} \mid \mathbf{D}) =$

$$\text{posterior} \propto \text{likelihood} \times \text{prior}$$

PAUSE

Introduction to statistics