Basic Statistics

Basic Statistics

- Deep learning requires understanding basic Statistics
  - Introduction of notation
  - Review of basic concepts

Probability: An Example

- Toss a coin. Will it land on heads or tails?

Dynamical Bias in the Coin Toss, Diaconis etal 2007 ↩️

Basic Notation

$$P(X = a)$$

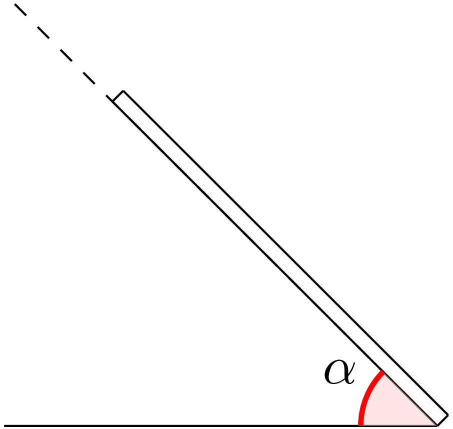
Basic Notation: An Example

$$P(\overbrace{\boxed{\mathrm{X}} = \boxed{\mathrm{a}}}^{\text{an event}})$$

a random variable ⌟ ⌞ a specific value

Probability Density

- What is $P(Y = \alpha)$?
  - not defined
- Cumulative probability: $P(\alpha_1 \leq Y < \alpha_2)$
- Probability density: $p(Y = \alpha)$
  - $$p(Y = \alpha) = \frac{P(\alpha - \epsilon \leq Y < \alpha + \epsilon)}{2\epsilon}$$



Unified Notation

Symbol	Description
$P(x) = P(X = x)$	Probability of a discrete event
$P(x) = p(X = x)$	Probability density of a continuous event

Properties of P(x)

A function $P(x)$ that captures the probabilities of any value x.

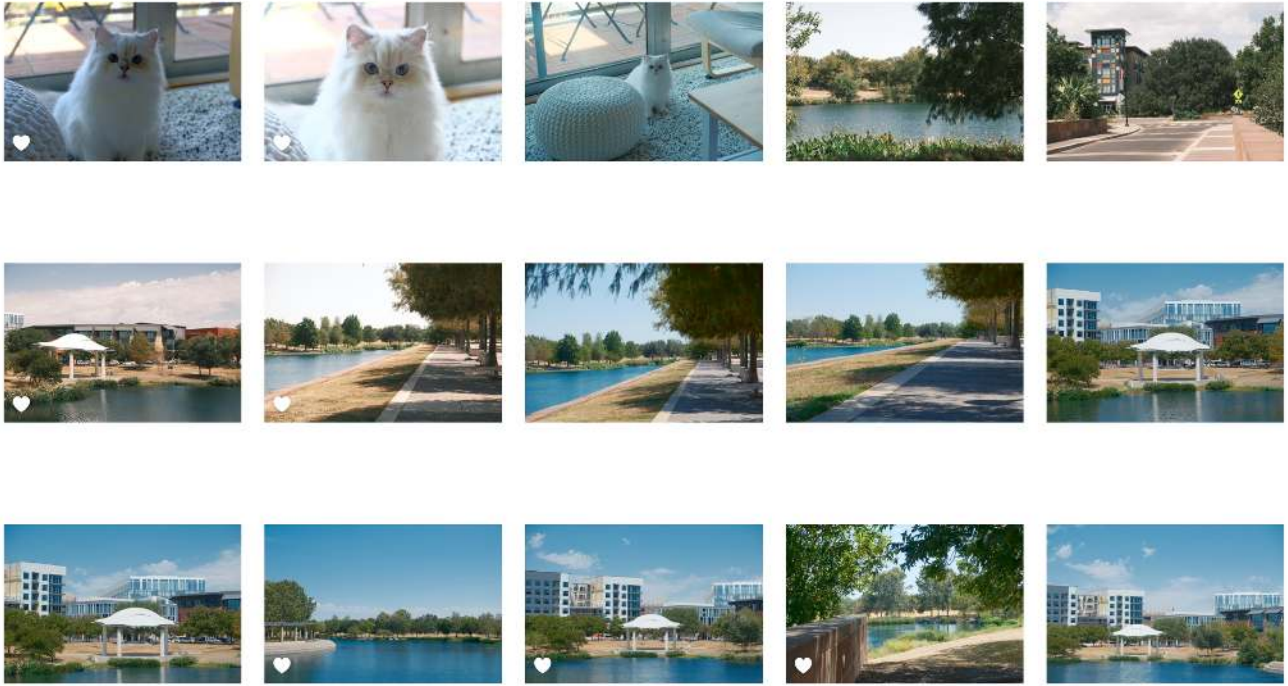
Property	
Non-negativity	$0 \leq P(x)$
Boundedness (discrete only)	$0 \leq P(x) \leq 1$
Summation (discrete)	$E_P[1] = \sum_x P(x) = 1$
Summation (continuous)	$E_P[1] = \int P(x) dx = 1$

Probability Distributions

$$P(x) = P(\overbrace{\underset{\text{a random variable}}{\boxed{\text{X}}} = \underset{\text{a specific value}}{\boxed{\text{x}}}}^{\text{an event}})$$

- What is P?
  - A function
  - Discrete $P : \{c_1, c_2, \ldots, c_n\} \to [0, 1]$
  - Continuous $P : \mathbb{R} \to \mathbb{R}$
  - Called a **probability distribution**

Examples of Probability Distributions: Visual World



Examples of Probability Distributions: Weather

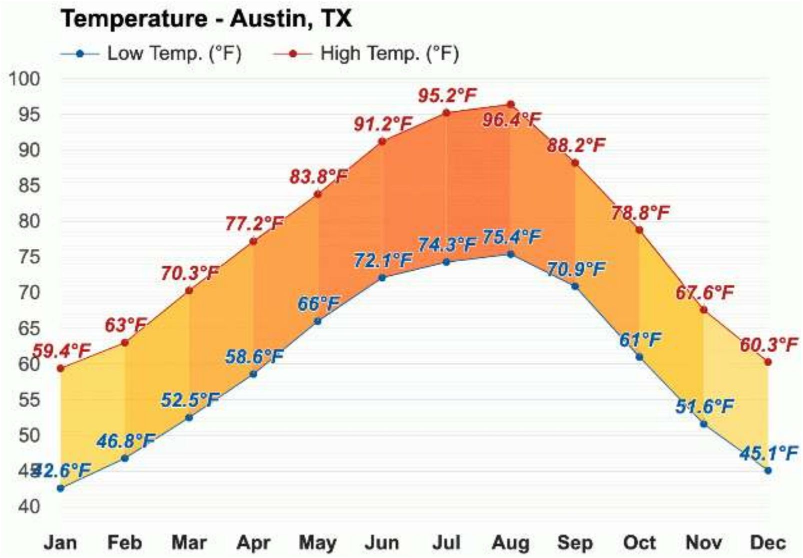
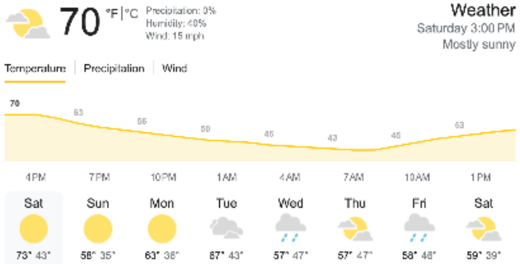


Image source: Weather U.S.

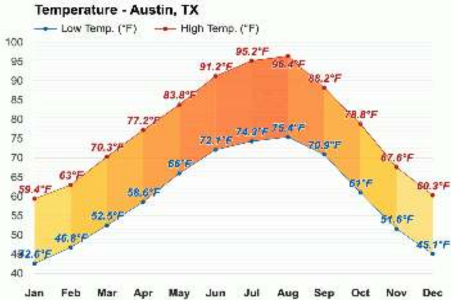
Three Types of Distribution

Data Generating

Empirical



Model



Basic Statistic Concepts

Expectation

$$E_{x \sim P}\left[f(x)\right] = \begin{cases} \sum_x P(x) f(x) & \text{discrete } P \\ \\ \int_x P(x) f(x) dx & \text{continuous } P \end{cases}$$

Expectation - Short Forms

- $E_P\left[f(x)\right]$
- $E\left[f(x)\right]$
- $E_P\left[f\right]$
- $E\left[f\right]$

Linearity of Expectation

$$E\left[f(x) + g(x)\right] = E\left[f(x)\right] + E\left[g(x)\right]$$

$$E\left[\alpha f(x)\right] = \alpha E\left[f(x)\right]$$

Mean and Variance

Mean:

$$\mu_x = E_{x \sim P}\left[x\right]$$

Variance:

$$\sigma_x^2 = Var_{x \sim P}\left[x\right] = E_{x \sim P}\left[(x - \mu_x)^2\right]$$

Unified Notations: A Word of Warning

Discrete distribution	**Continuous distribution**
$Var_{x \sim P}\left[x\right]$ is always finite	$Var_{x \sim P}\left[x\right]$ can be infinite
P(X) always less than 1	P(X) can be larger than 1
…	…

Sampling

$$x \sim P$$

Example of Sampling

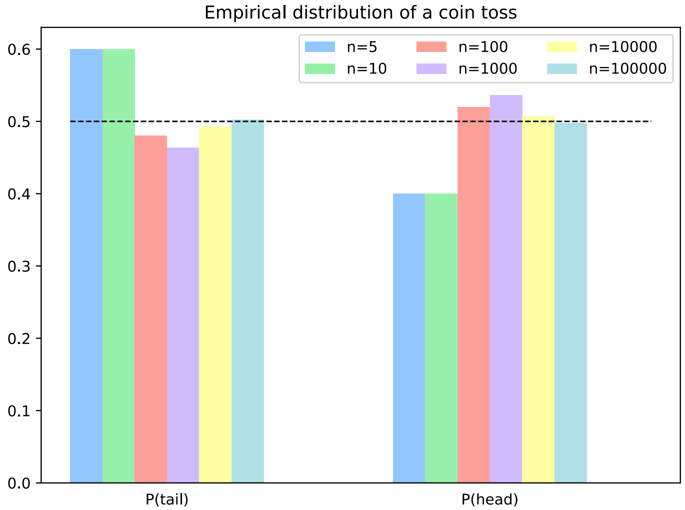
Example of Sampling

Example of Sampling

Bias in Samples

- Samples are always biased
- For infinite samples:

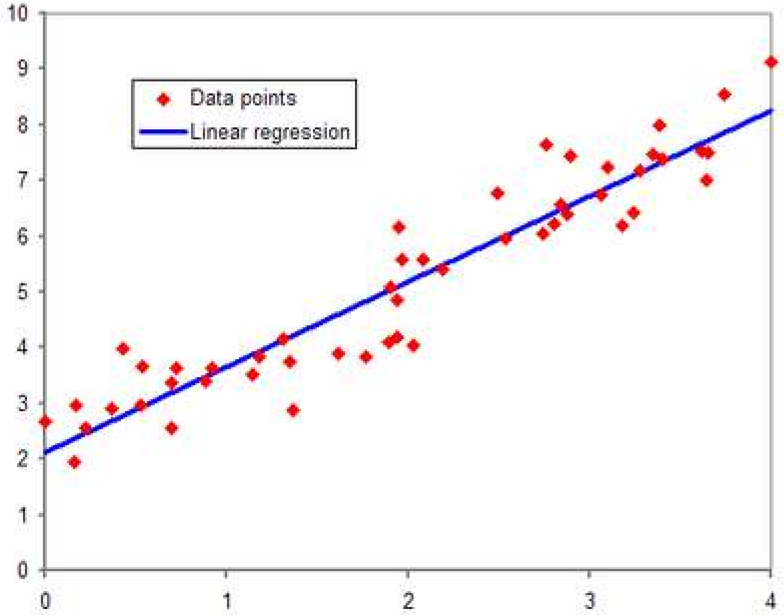
  empirical distribution = data generating distribution



Statistical Models

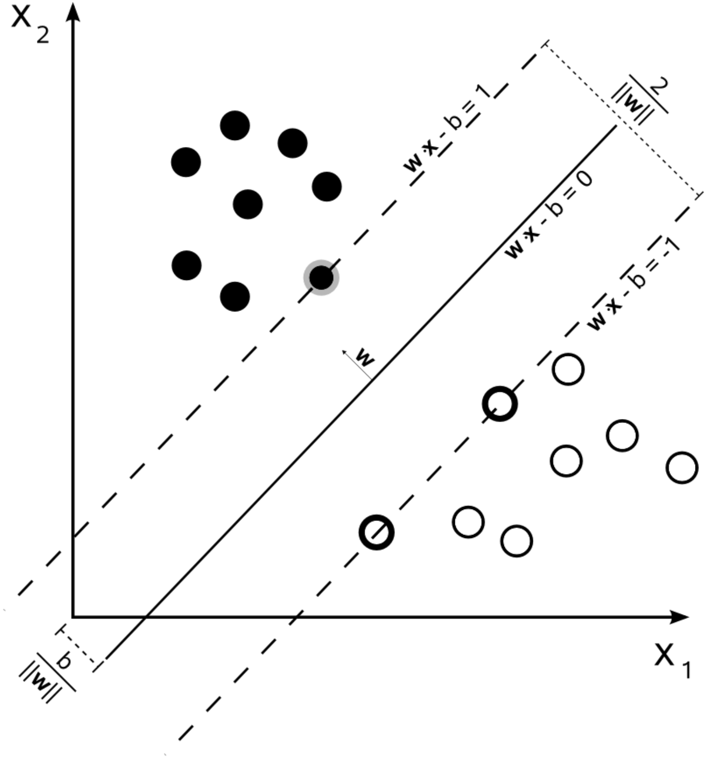
Regression Model

$$f_\theta : \mathbb{R}^n \to \mathbb{R}^d$$



Classification Model

$$f_\theta : \mathbb{R}^n \to P(X), P(X) \subset \mathbb{R}^d$$



Statistical Model Summary

$$f_\theta : X \to Y$$

- X: input space
- Y: output space
- θ: model parameters

In machine learning:

- Goal: find the optimal **parameter** θ
- How: learn from **data**

Data

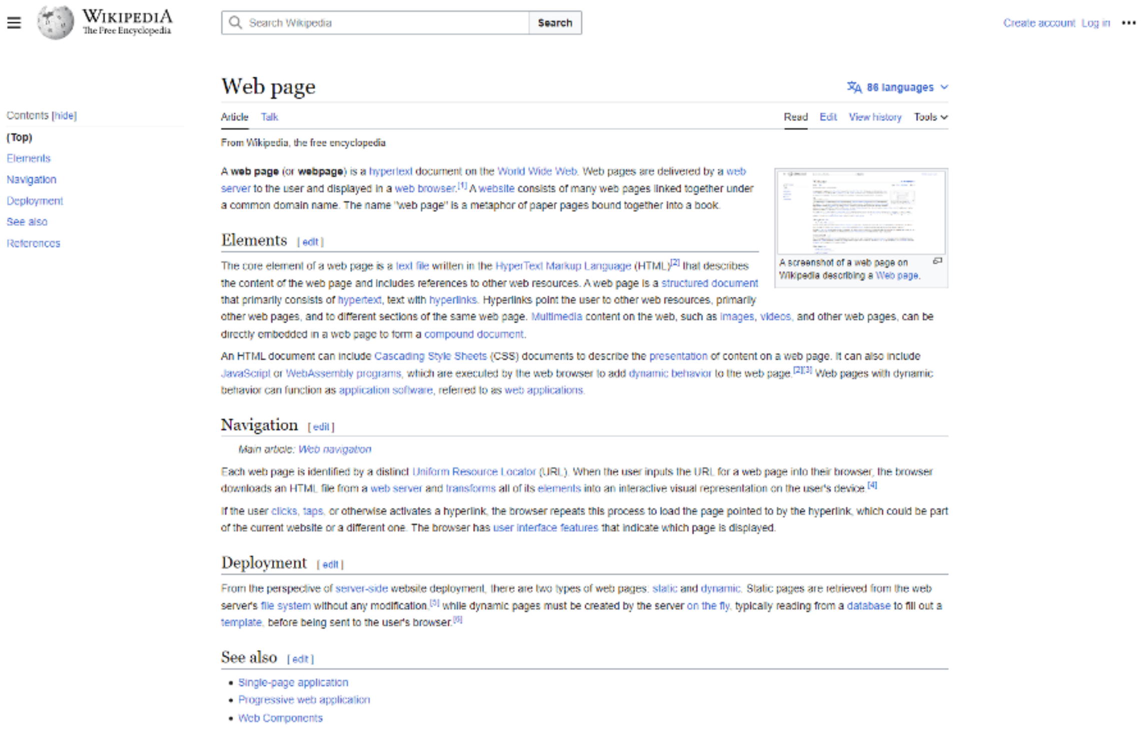
Unlabeled Data:

$$D = \{x_1, x_2, \ldots\} \qquad \text{where } x_i \sim P_D.$$

Labeled Data:

$$D = \{(x_1, l_1), (x_2, l_2), \ldots\} \qquad \text{where } (x_i, l_i) \sim P_D,$$

Examples of Data: Internet



Web page

From Wikipedia, the free encyclopedia

A **web page** (or **webpage**) is a hypertext document on the World Wide Web. Web pages are delivered by a web server to the user and displayed in a web browser.[1] A website consists of many web pages linked together under a common domain name. The name "web page" is a metaphor of paper pages bound together into a book.

A screenshot of a web page on Wikipedia describing a Web page.

Elements [edit]

The core element of a web page is a text file written in the HyperText Markup Language (HTML)[2] that describes the content of the web page and includes references to other web resources. A web page is a structured document that primarily consists of hypertext, text with hyperlinks. Hyperlinks point the user to other web resources, primarily other web pages, and to different sections of the same web page. Multimedia content on the web, such as images, videos, and other web pages, can be directly embedded in a web page to form a compound document.

An HTML document can include Cascading Style Sheets (CSS) documents to describe the presentation of content on a web page. It can also include JavaScript or WebAssembly programs, which are executed by the web browser to add dynamic behavior to the web page.[2][3] Web pages with dynamic behavior can function as application software, referred to as web applications.

Navigation [edit]

Main article: Web navigation

Each web page is identified by a distinct Uniform Resource Locator (URL). When the user inputs the URL for a web page into their browser, the browser downloads an HTML file from a web server and transforms all of its elements into an interactive visual representation on the user's device.[4]

If the user clicks, taps, or otherwise activates a hyperlink, the browser repeats this process to load the page pointed to by the hyperlink, which could be part of the current website or a different one. The browser has user interface features that indicate which page is displayed.

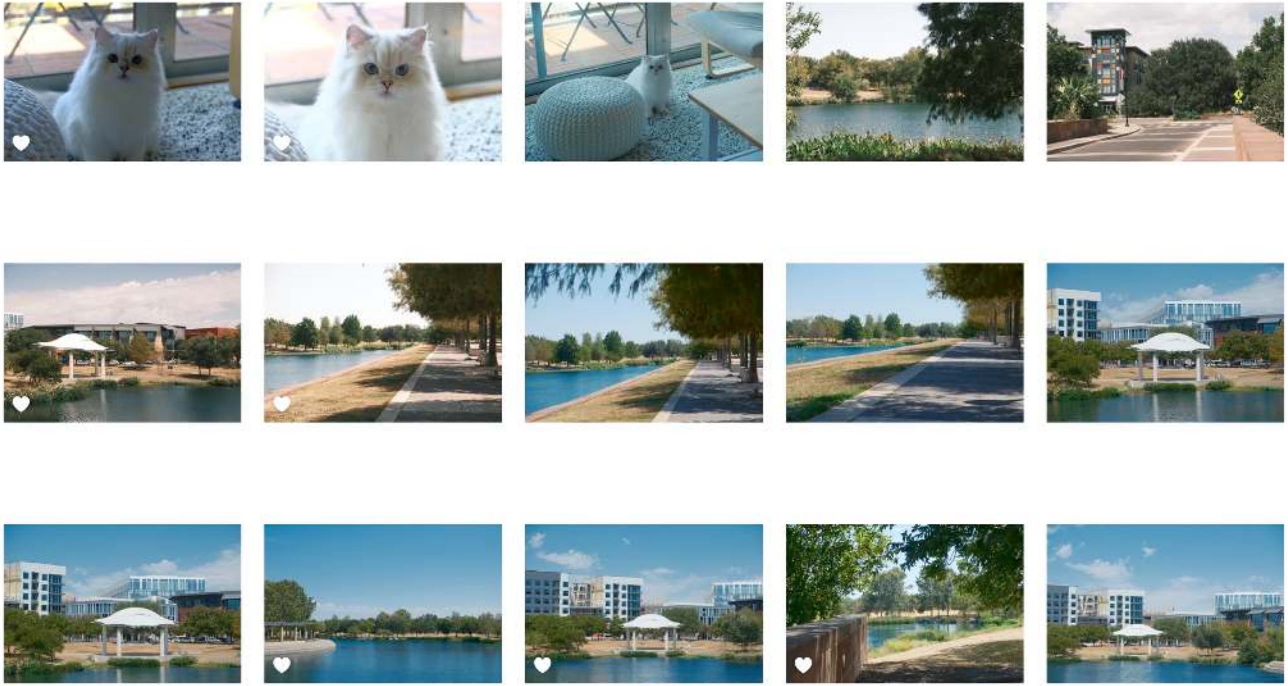
Deployment [edit]

From the perspective of server-side website deployment, there are two types of web pages: static and dynamic. Static pages are retrieved from the web server's file system without any modification,[5] while dynamic pages must be created by the server on the fly, typically reading from a database to fill out a template, before being sent to the user's browser.[6]

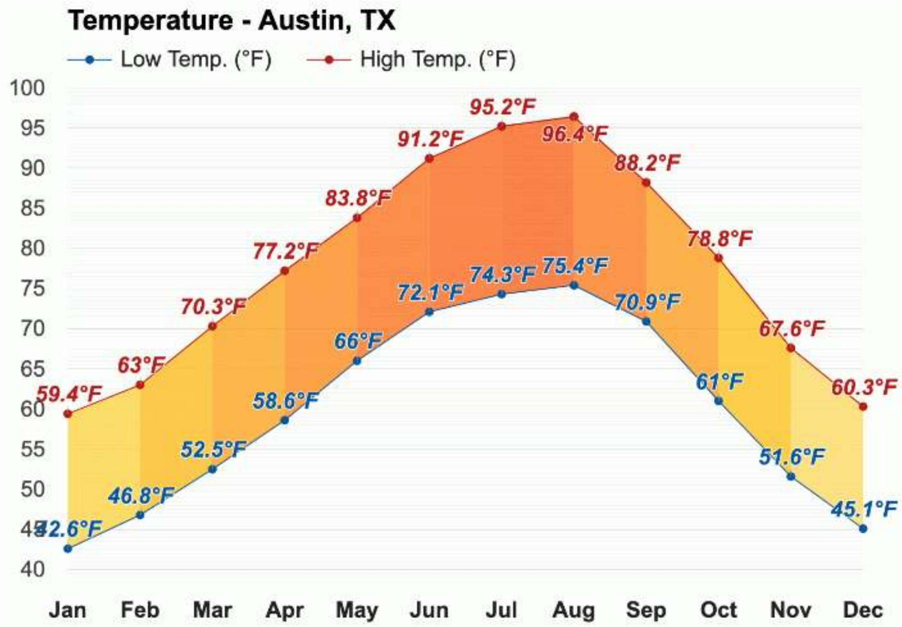
See also [edit]

- Single-page application
- Progressive web application
- Web Components

Examples of Data: Personal Photo Collection



Examples of Data: Weather



Source: Weather U.S.

Basic Statistics - TL;DR

Model: a function of the data to predict values $f_\theta : X \to Y$

Expectation: measures the value of f weighted by probability P.

Variance: measures deviation from the mean