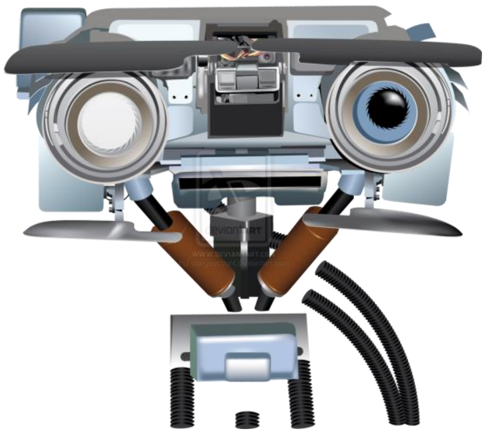
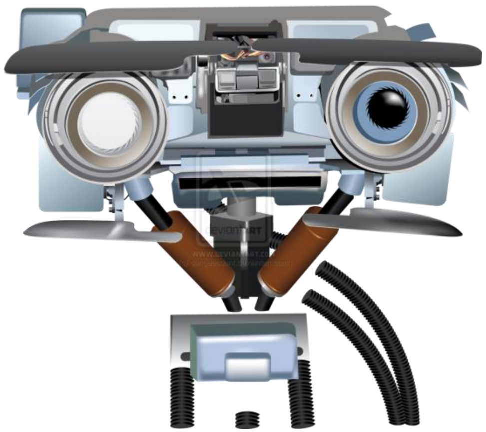


Preliminaries

Kate Saenko

Today

- Recap of video "readings"
- Review of probability theory
  - Sum Rule
  - Product Rule
  - Bayes Rule
  - Bayesian vs Frequentist Probability
- Review of matrices and vectors
  - Sum, product, transpose, etc.



Announcements

Check "Schedule" often!

- Psets posted to Schedule page
- Also readings, video readings
- The lectures will be captured on video and posted online
- Quiz0: see me after class if you have not handed it in

Problem Sets

ps1 will be released today

- Two components
  - Programming
    - you will submit code, write up and your grade
    - we will randomly check code to verify
  - Theoretical
    - practice mathematical and conceptual knowledge
    - we will randomly pick one problem to grade
- 6 psets
  - roughly every two weeks
  - should take 5-10 hours each

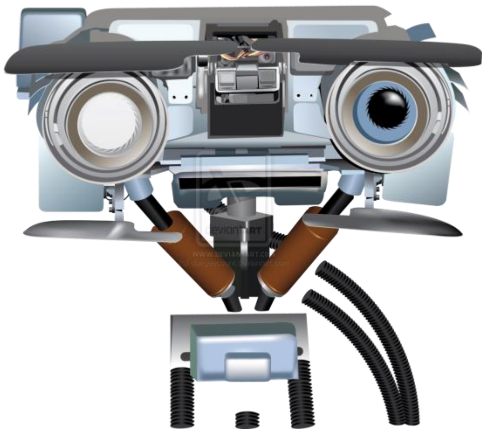
Assigned video "reading"

I. Introduction (Week 1)

- Welcome (7 min)
- What is Machine Learning? (7 min)
- Supervised Learning (12 min)
- Unsupervised Learning (14 min)

III. Linear Algebra Review (Week 1, Optional)

- Matrices and Vectors (9 min)
- Addition and Scalar Multiplication (7 min)
- Matrix Vector Multiplication (14 min)
- Matrix Matrix Multiplication (11 min)
- Matrix Multiplication Properties (9 min)
- Inverse and Transpose (11 min)

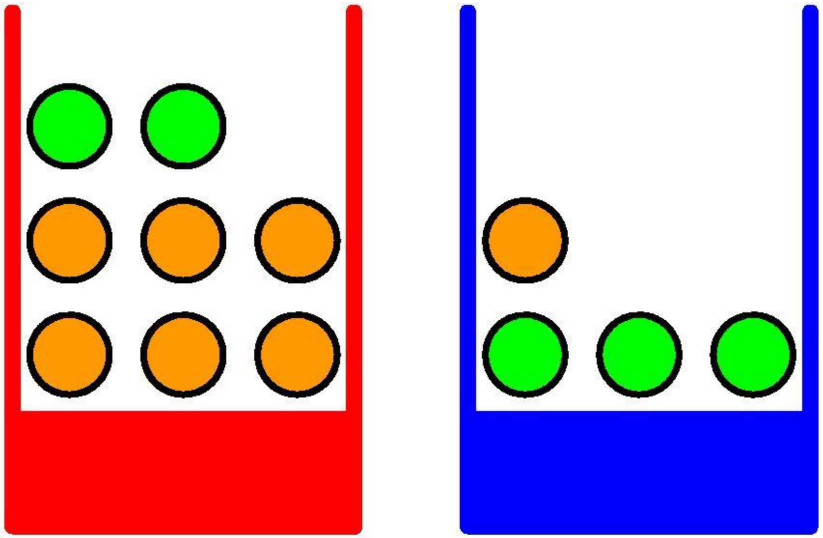


Probability Theory Review

A simple example

Probability Theory

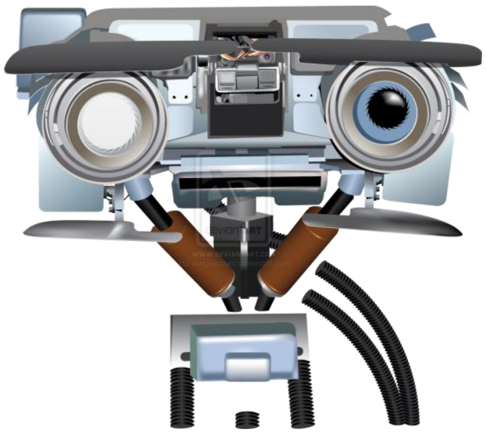
see Bishop Chapter 1.2



- Pick a random box
- Pick a random fruit
- Observe the fruit type (orange or apple)
- Put it back in the box
- Repeat trial many times

What is the probability of picking an apple?

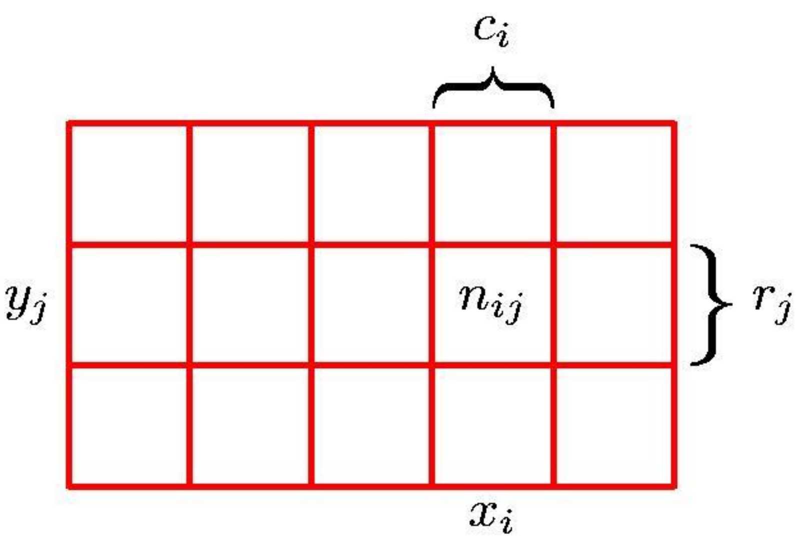
C. Bishop



Probability Theory Review

Useful Definitions

Probability Theory



Marginal Probability

$$p(X = x_i) = \frac{c_i}{N}.$$

Joint Probability

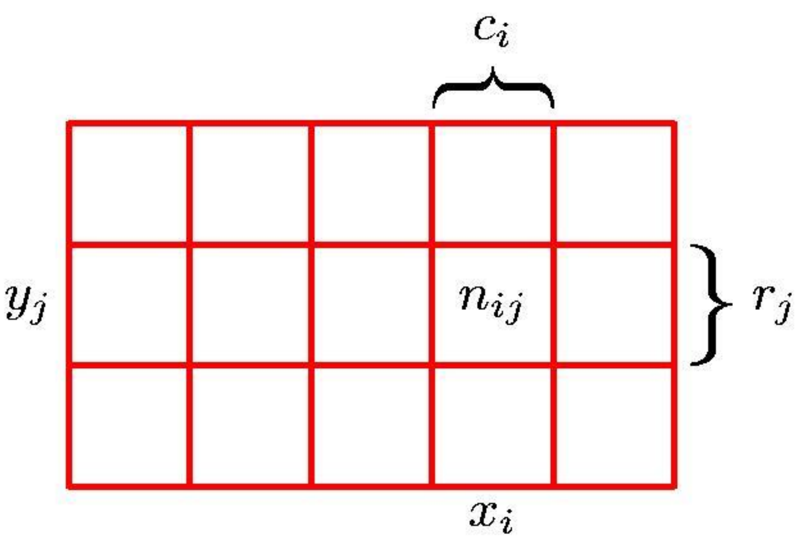
$$p(X = x_i, Y = y_j) = \frac{n_{ij}}{N}$$

Conditional Probability

$$p(Y = y_j | X = x_i) = \frac{n_{ij}}{c_i}$$

C. Bishop

Probability Theory



Sum Rule

$$p(X = x_i) = \frac{c_i}{N} = \frac{1}{N}\sum_{j=1}^{L} n_{ij}$$

$$= \sum_{j=1}^{L} p(X = x_i, Y = y_j)$$

Product Rule

$$\begin{aligned} p(X = x_i, Y = y_j) &= \frac{n_{ij}}{N} = \frac{n_{ij}}{c_i} \cdot \frac{c_i}{N} \\ &= p(Y = y_j | X = x_i) p(X = x_i) \end{aligned}$$

C. Bishop

The Rules of Probability

Sum Rule $$p(X) = \sum_Y p(X, Y)$$

Product Rule $$p(X, Y) = p(Y|X)p(X)$$

C. Bishop

Bayes' Theorem

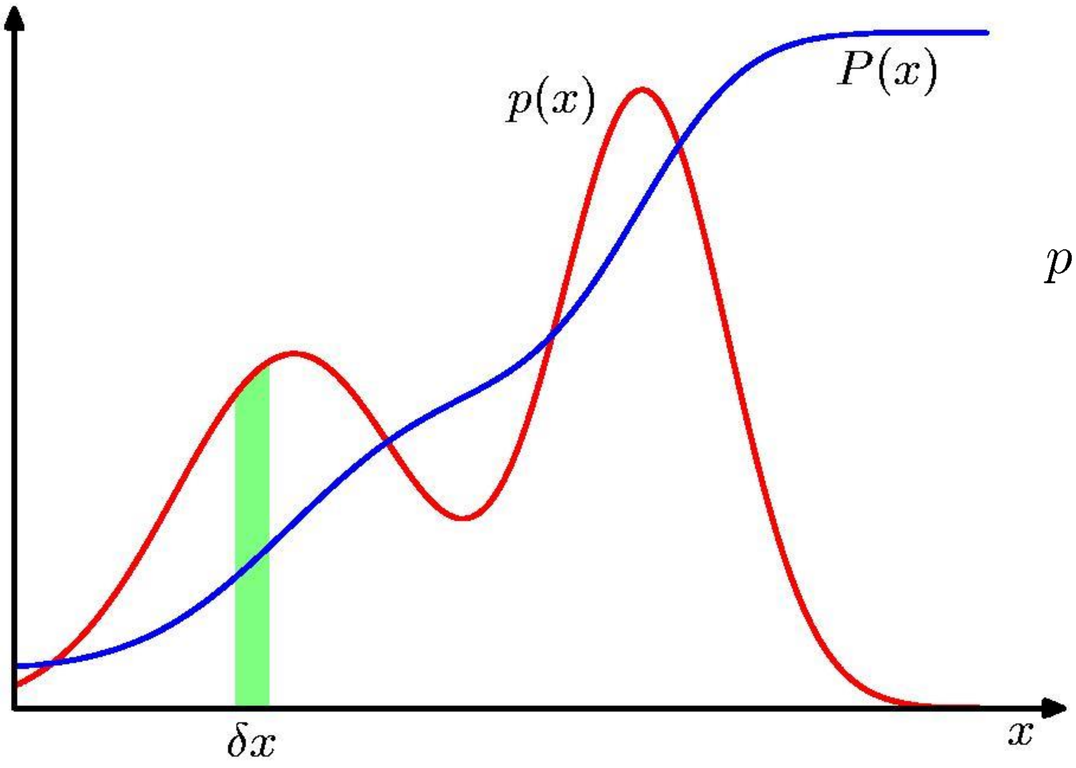
$$p(Y|X) = \frac{p(X|Y)p(Y)}{p(X)}$$

$$p(X) = \sum_Y p(X|Y)p(Y)$$

posterior $\propto$ likelihood × prior

C. Bishop

Probability Densities



$$p(x \in (a,b)) = \int_a^b p(x)\,\mathrm{d}x$$

$$P(z) = \int_{-\infty}^{z} p(x)\,\mathrm{d}x$$

$$p(x) \geqslant 0 \qquad \int_{-\infty}^{\infty} p(x)\,\mathrm{d}x = 1$$

C. Bishop

Expectations

$$\mathbb{E}[f] = \sum_x p(x) f(x)$$

$$\mathbb{E}[f] = \int p(x) f(x) \, \mathrm{d}x$$

$$\mathbb{E}_x[f|y] = \sum_x p(x|y) f(x)$$

Conditional Expectation
(discrete)

$$\mathbb{E}[f] \simeq \frac{1}{N} \sum_{n=1}^{N} f(x_n)$$

Approximate Expectation
(discrete and continuous)

C. Bishop

Variances and Co-variances

$$\mathrm{var}[f] = \mathbb{E}\left[\left(f(x) - \mathbb{E}[f(x)]\right)^2\right] = \mathbb{E}[f(x)^2] - \mathbb{E}[f(x)]^2$$

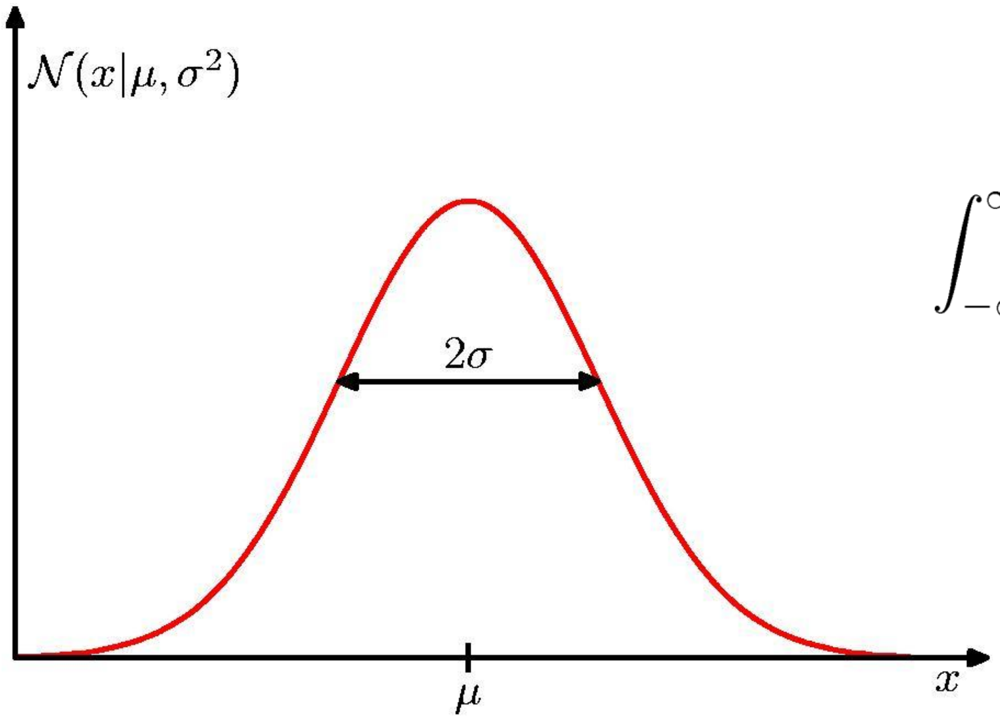
$$\begin{aligned} \mathrm{cov}[x, y] &= \mathbb{E}_{x,y}\left[\{x - \mathbb{E}[x]\}\{y - \mathbb{E}[y]\}\right] \\ &= \mathbb{E}_{x,y}[xy] - \mathbb{E}[x]\mathbb{E}[y] \end{aligned}$$

$$\begin{aligned} \mathrm{cov}[\mathbf{x}, \mathbf{y}] &= \mathbb{E}_{\mathbf{x},\mathbf{y}}\left[\{\mathbf{x} - \mathbb{E}[\mathbf{x}]\}\{\mathbf{y}^{\mathrm{T}} - \mathbb{E}[\mathbf{y}^{\mathrm{T}}]\}\right] \\ &= \mathbb{E}_{\mathbf{x},\mathbf{y}}[\mathbf{x}\mathbf{y}^{\mathrm{T}}] - \mathbb{E}[\mathbf{x}]\mathbb{E}[\mathbf{y}^{\mathrm{T}}] \end{aligned}$$

C. Bishop

The Gaussian Distribution

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



$$\mathcal{N}(x|\mu,\sigma^2) > 0$$

$$\int_{-\infty}^{\infty} \mathcal{N}\left(x|\mu,\sigma^2\right)\,\mathrm{d}x = 1$$

C. Bishop

Gaussian Mean and Variance

$$\mathbb{E}[x] = \int_{-\infty}^{\infty} \mathcal{N}\left(x|\mu, \sigma^2\right) x \,\mathrm{d}x = \mu$$

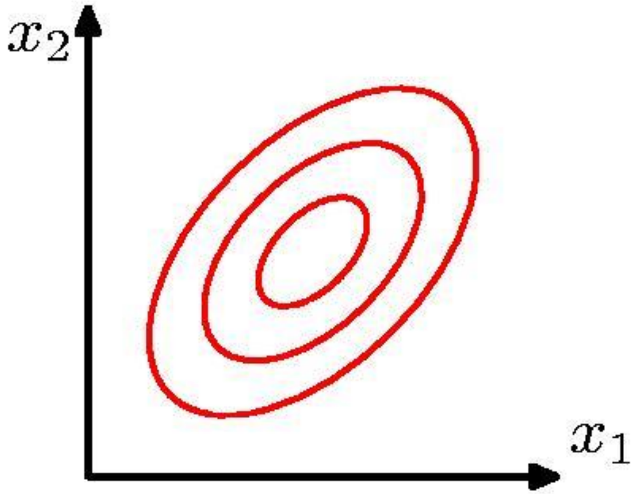
$$\mathbb{E}[x^2] = \int_{-\infty}^{\infty} \mathcal{N}\left(x|\mu, \sigma^2\right) x^2 \,\mathrm{d}x = \mu^2 + \sigma^2$$

$$\mathrm{var}[x] = \mathbb{E}[x^2] - \mathbb{E}[x]^2 = \sigma^2$$

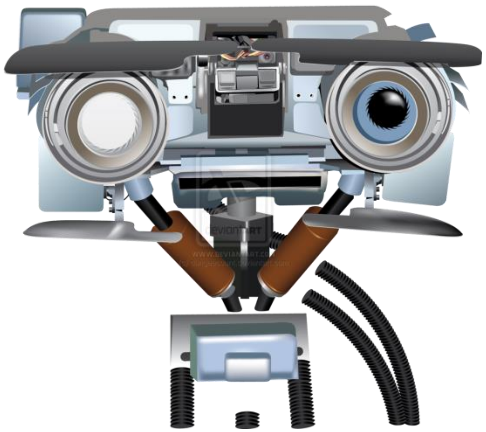
C. Bishop

The Multivariate Gaussian

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



C. Bishop



Linear Algebra review

Matrices and vectors

Matrix Elements (entries of matrix)

$$A = \begin{bmatrix} 1402 & 191 \\ 1371 & 821 \\ 949 & 1437 \\ 147 & 1448 \end{bmatrix}$$

$A_{ij} =$ "i, j entry" in the i^{th} row, j^{th} column.

Vector: An n x 1 matrix.

$$y = \begin{bmatrix} 460 \\ 232 \\ 315 \\ 178 \end{bmatrix}$$

$y_i = i^{th}$ element

1-indexed vs 0-indexed:

$$y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} \qquad y = \begin{bmatrix} y_0 \\ y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

Matrix Addition

$$\begin{bmatrix} 1 & 0 \\ 2 & 5 \\ 3 & 1 \end{bmatrix} + \begin{bmatrix} 4 & 0.5 \\ 2 & 5 \\ 0 & 1 \end{bmatrix} =$$

$$\begin{bmatrix} 1 & 0 \\ 2 & 5 \\ 3 & 1 \end{bmatrix} + \begin{bmatrix} 4 & 0.5 \\ 2 & 5 \end{bmatrix} =$$

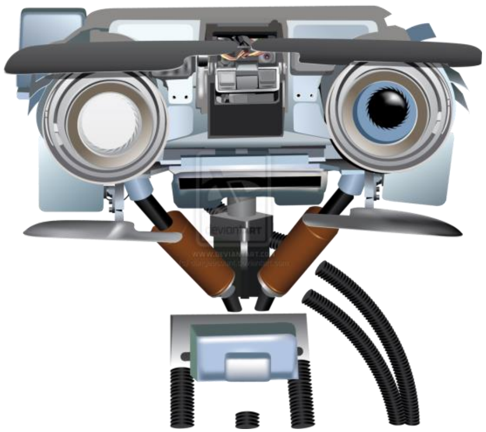
Scalar Multiplication

$$3 \times \begin{bmatrix} 1 & 0 \\ 2 & 5 \\ 3 & 1 \end{bmatrix} =$$

$$\begin{bmatrix} 4 & 0 \\ 6 & 3 \end{bmatrix} / 4 =$$

Combination of Operands

$$3 \times \begin{bmatrix} 1 \\ 4 \\ 2 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 5 \end{bmatrix} - \begin{bmatrix} 3 \\ 0 \\ 2 \end{bmatrix} /3$$



Linear Algebra review

Matrix-vector multiplication

Example

$$\begin{bmatrix} 1 & 3 \\ 4 & 0 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 5 \end{bmatrix} =$$

Details:

$$A \quad \times \quad x \quad = \quad y$$

$$\begin{bmatrix} \quad & \quad & \quad \\ \\ \\ \end{bmatrix} \times \begin{bmatrix} \quad \\ \\ \\ \\ \end{bmatrix} = \begin{bmatrix} \quad \\ \\ \\ \end{bmatrix}$$

m x n matrix (m rows,n columns)	n x 1 matrix (n-dimensional vector)	m-dimensional vector

To get y_i, multiply A's i^{th} row with elements of vector x, and add them up.

Example

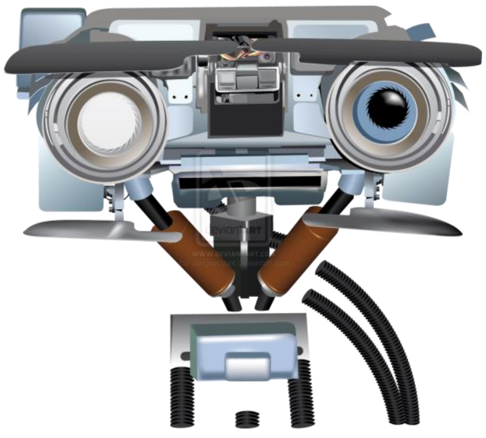
$$\begin{bmatrix} 1 & 2 & 1 & 5 \\ 0 & 3 & 0 & 4 \\ -1 & -2 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \\ 2 \\ 1 \end{bmatrix} =$$

House sizes:

$$\begin{matrix} 2104 \\ 1416 \\ 1534 \\ 852 \end{matrix}$$

$$h_\theta(x) = -40 + 0.25x$$

How do we get predicted price as matrix-vector product?



Linear Algebra review

Matrix-matrix multiplication

Example

$$\begin{bmatrix} 1 & 3 & 2 \\ 4 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 0 & 1 \\ 5 & 2 \end{bmatrix} =$$

$$\begin{bmatrix} 1 & 3 & 2 \\ 4 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} 1 \\ 0 \\ 5 \end{bmatrix} =$$

$$\begin{bmatrix} 1 & 3 & 2 \\ 4 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix} =$$

Details:

$$A \times B = C$$

$$\begin{bmatrix} \quad \end{bmatrix} \times \begin{bmatrix} \quad \end{bmatrix} = \begin{bmatrix} \quad \end{bmatrix}$$

m x n matrix (m rows, n columns)	n x o matrix (n rows, o columns)	m x o matrix

The i^{th} column of the matrix C is obtained by multiplying A with the i^{th} column of B. (for i = 1,2,…,o)

Given house sizes:

$$\begin{matrix} 2104 \\ 1416 \\ 1534 \\ 852 \end{matrix}$$

What is the price of each house?

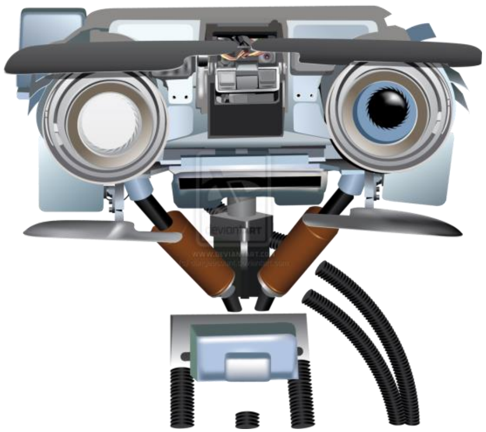
Have 3 competing linear functions:

1. $h_\theta(x) = -40 + 0.25x$
2. $h_\theta(x) = 200 + 0.1x$
3. $h_\theta(x) = -150 + 0.4x$

Matrix

Matrix

$$\begin{bmatrix} 1 & 2104 \\ 1 & 1416 \\ 1 & 1534 \\ 1 & 852 \end{bmatrix} \times \begin{bmatrix} -40 & 200 & -150 \\ 0.25 & 0.1 & 0.4 \end{bmatrix} = \begin{bmatrix} 486 & 410 & 692 \\ 314 & 342 & 416 \\ 344 & 353 & 464 \\ 173 & 285 & 191 \end{bmatrix}$$



Linear Algebra Review

Matrix multiplication properties

Let A and B be matrices. Then in general,

$A \times B \neq B \times A$. (not commutative.)

E.g. $$\begin{bmatrix}1 & 1\\0 & 0\end{bmatrix}\begin{bmatrix}0 & 0\\2 & 0\end{bmatrix} = \begin{bmatrix}2 & 0\\0 & 0\end{bmatrix}$$

$$\begin{bmatrix}0 & 0\\2 & 0\end{bmatrix}\begin{bmatrix}1 & 1\\0 & 0\end{bmatrix} = \begin{bmatrix}0 & 0\\2 & 2\end{bmatrix}$$

Associative

$$A \times B \times C.$$

Let $D = B \times C.$ Compute $A \times D.$

Let $E = A \times B.$ Compute $E \times C.$

Identity Matrix

Denoted I (or $I_{n \times n}$).
Examples of identity matrices:

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$
2 x 2

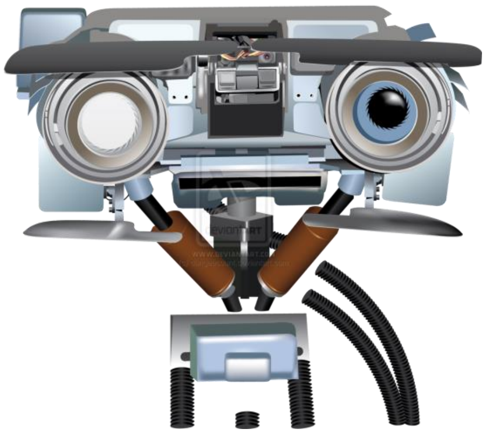
$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$
3 x 3

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$
4 x 4

For any matrix A,

$$A \cdot I = I \cdot A = A$$

In general, is AB = BA?



Linear Algebra review

Inverse and transpose

Not all numbers have an inverse

Matrix inverse:

If A is an m x m matrix, and if it has an inverse,

$$AA^{-1} = A^{-1}A = I.$$

For a 2 x 2 matrix, what is a sufficient condition for it to have an inverse?

Matrices that don't have an inverse are "singular" or "degenerate"

Matrix Transpose

Example:

$$A = \begin{bmatrix} 1 & 2 & 0 \\ 3 & 5 & 9 \end{bmatrix} \qquad A^T = \begin{bmatrix} 1 & 3 \\ 2 & 5 \\ 0 & 9 \end{bmatrix}$$

Let A be an m x n matrix, and let $B = A^T$.
Then B is an n x m matrix, and

$$B_{ij} = A_{ji}.$$

Next Lecture

- Linear Algebra basics: solution of linear systems, etc.
- Convex optimization