1. Intro. + Math + Modelling

Linear Systems

Linear Algebra Table

Scalar $c \in \mathbb{R}$	**Vector** $b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} \in \mathbb{R}^n$	**Matrix** $A = \begin{bmatrix} A_{11} & A_{12} & \dots \\ A_{21} & \ddots & \\ \vdots & & A_{nm} \end{bmatrix} \in \mathbb{R}^{n \times m}$ - Fat matrix: n<m - Skinny matrix: n>m
Matrix Transpose $A^T = \begin{bmatrix} A_{11} & A_{21} & \dots \\ A_{12} & \ddots & \\ \vdots & & A_{mn} \end{bmatrix}$	**Matrix Addition** $A + B = \begin{bmatrix} A_{11} + B_{11} & A_{12} + B_{12} & \dots \\ A_{21} + B_{21} & \ddots & \\ \vdots & & A_{mn} + B_{mn} \end{bmatrix}$	**Matrix Multiplication** $AB = \begin{bmatrix} \sum_i A_{1i}B_{i1} & \sum_i A_{1i}B_{i2} & \dots \\ \sum_i A_{2i}B_{i1} & \ddots & \\ \vdots & & \sum_i A_{mi}B_{in} \end{bmatrix}$
Matrix transpose of Added Matrices $(A+B)^T = A^T + B^T$	**Matrix Transpose of Multiplied Matrices** $(AB)^T = B^T A^T$	**Quadratic Form** $(Ax+b)^T(Ax+b) = x^TA^TAx + 2x^TA^Tv + b^Tb$ $= \underbrace{x^TCx}_{\text{quadratic term}} + d^Tx + e$
Matrix Rank: $\rho(A)$ - The number of independent rows or columns - **Nonsingular =** Full Rank : $\rho(A) = \min(n, m)$ - **Singular =** Not Full Rank: $\rho(A) < \min(n, m)$ -- Non-empty nullspace: $\exists x$ such that $Ax = 0$	**Matrix Inverse (square A)** $AA^{-1} = A^{-1}A = I$ - **Nonsingular (Full Rank) and square** **. => INVERTIBLE**	**Symmetric Matrix** $A = A^T = \begin{bmatrix} A_{11} & A_{12} & \dots \\ A_{21} & \ddots & \\ \vdots & & A_{nm} \end{bmatrix}$
Matrix Trace $tr(A) = \sum_i A_{ii}$	**Positive Definiteness (Semi-Definiteness)** - For a symmetric $n \times n$ matrix A, and for any x in $\mathbb{R}^n$: $x^TAx > 0 \qquad (x^TAx \geq 0)$	**Eigenvalues and Eigenvectors** of a matrix - For a matrix A, --- the vector x is an *eigenvector* of A --- with a corresponding *eigenvalue* λ -- if they satisify the equation: $Ax = \lambda x$ NOTE: - The eigenvalues of a diagonal matrix are its diagonal elements - The inverse of A exists iff none of the eigenvalues are zero --- $(\lambda_i \neq 0)$ - Positive definite A has all eigenvalues greater than zero --- $(\lambda_i > 0 \, \forall i)$
Differentiation of Linear Matrix Equation $\frac{d}{dx}(Ax) = A$ $\frac{d}{dx}(A^Tx) = A^T$	**Differentiation of a quadratic matrix equation** $\frac{d}{dx}(x^TAx) = x^TA + x^TA^T$	

Least Squares Solution

- If A is a skinny matrix (n>m), and we wish to find x for which $Ax = b$
- Since A is skinny, the problem is **over-constrained** => **No solution exists**
- Instead, minimize the square of the error between Ax and b:
    -

$$
\begin{aligned}
\min_x ||Ax - b||_2^2 \\
&= \min_x (Ax-b)^T(Ax-b) \\
&= \min_x x^T A^T A x - 2b^T A x + b^T b
\end{aligned}
$$

- Setting the derivative to zero
    - $$
\begin{aligned}
2x^T A^T A - 2b^T A = 0 \\
A^T A x &= A^T b \\
x &= (A^T A)^{-1} A^T b \\
x = A^{\dagger} b
\end{aligned}
$$
    - $A^{\dagger}$: **Pseudo-inverse**
- This terminology is used over and over, Quadratic cost minimized to find closed form solution

---

Probability

Probability Algebra Table

Measures of Distributions $\mathcal{N}(\mu, \sigma^2)$ **Mean:** - Expected value of a random variable: $\mu = E[x]$ -- Discrete Case: $\mu = \sum_{i=1}^{n} x_i p(x_i)$ -- Continuous Case: $\mu = \int x p(x) dx$ **Variance:** - Measure of the variability of a random variable: $\sigma^2(x) = E[(x-\mu)^2]$ -- Discrete Case: $\sigma^2(x) = \sum_{i=1}^{n} (x_i - \mu)^2 p(x_i)$ -- Continuous Case: $\sigma^2(x) = \int (x-\mu)^2 p(x) dx$	**Multi-variable Distributions** **Mean:** $\mu = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix} = \begin{bmatrix} E[X_1] \\ \vdots \\ E[X_n] \end{bmatrix}$ **Covariance:** - Measure of how much two random variables change together: . $Cov(X_i, X_j) = E[(X_i - \mu_i)(X_j - \mu_j)] = E[X_i X_j] - \mu_i \mu_j$ -- If $Cov(X, Y) > 0$, when X is above its expected values, then Y tends to be above its expected value -- If $Cov(X, Y) < 0$, when X is above its expected value, then Y tends to be below its expected value -- If X, Y are independent, **Covariance Matrix,** Σ - Defins variational relationship between each pair of random variables: $\Sigma_{i,j} = Cov(X_i, X_j)$ - Generalization of variance, diagonal elements represent variance of each random variable: $Cov(X_i, X_i) = Var(X_i)$ - Covariance matrix is **Symmetric, Positive semi-definite**
Multiplication by a constant matrix yeilds	**Addition/ Subtraction of random variables**

$$
\begin{aligned}
cov(Ax) &= E[(Ax - A\mu)(Ax - A\mu)^T] \\
&= E[A(x-\mu)(x-\mu)^T A^T] \\
&= AE[(x-\mu)(x-\mu)^T]A^T \\
&= A\,cov(x)\,A^T
\end{aligned}
$$

$cov(X \pm Y) = cov(X) + cov(Y) \pm 2cov(X,Y)$

- If X,Y independent,
$cov(X \pm Y) = cov(X) + cov(Y)$

Joint Probability
- Probability of x and y:
$p(X = x \text{ and } Y = y) = p(x, y)$

Independence
- If X,Y are independent, then
$p(x,y) = p(x)p(y)$

Conditional Probability
- Probability of x given y:
$p(X = x|Y = y) = p(x|y)$

- Relation to joint probability:
$p(x|y) = \frac{p(x,y)}{p(y)}$
- If X and Y are independent,
$p(x|y) = p(x)$

Law of Total Probability

- **Discrete**:
$\sum_x p(x) = 1$
$p(x) = \sum_y p(x,y)$
$p(x) = \sum_y p(x|y)p(y)$

- **Continuous**:
$\int p(x)dx = 1$
$p(x) = \int p(x,y)dy$
$p(x) = \int p(x|y)p(y)dy$

Probability distribution
- It is possible to define a discrete probability distribution as a column vector

$$
p(X = x) = \begin{bmatrix} p(X = x_1) \\ \vdots \\ p(X = x_n) \end{bmatrix}
$$

- The conditional probability can then be a matrix:

$$
p(x|y) = \begin{bmatrix} p(X = x_1|y_1) & \dots & p(x_1|y_m) \\ \vdots & \ddots & \vdots \\ p(X = x_n|y_1) & \dots & p(x_n|y_m) \end{bmatrix}
$$

Discrete Random Variable
- And the Law of Total Probabilities becomes
$p(x) = \sum_y p(x|y)p(y) = p(x|y) \cdot p(y)$

- Note, each column of $p(x|y)$ must sum to 1
$\sum_x p(x|y) = \sum_x \frac{p(x,y)}{py} = \frac{\sum_x p(y,x)}{p(y)} = \frac{p(y)}{p(y)} = 1$
- => Relation of joint and conditional probabilities => Total Probability = 1

Bayes Theorem
- From definition of conditional probability
$p(x|y) = \frac{p(x,y)}{p(y)}, \quad p(y|x) = \frac{p(x,y)}{p(x)}$
$p(x|y)p(y) = p(x,y) = p(y|x)p(x)$

- Bayes Theorem Defines how to update one's beliefs about **X** ,
given a known (new) value of **y**
$p(x|y) = \frac{p(y|x)p(x)}{p(y)} = \frac{\text{likelihood} \cdot \text{prior}}{\text{evidence}}$

- If **Y** is a measurement and **X** is the current vehicle state,
--- Bayes Theorem can be used to update the state estimate given a new measurement
-- **Prior**: probabilities that the vehicle is in any of the possible states
-- **Likelihood**: probability of getting the measurement that occurred given every possible state is the true state
-- **Evidence**: probability of getting the specific measurement recorded

Gaussian Distribution - $p(x) \sim \mathcal{N}(\mu, \sigma^2)$

$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}$

Multivariable - $p(x) \sim \mathcal{N}(\mu, \Sigma)$

$p(x) = \frac{1}{\det(2\pi\Sigma)} e^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)}$

Linear Combination:
$x \sim \mathcal{N}(\mu, \Sigma), \quad y = Ax + B$
$y \sim \mathcal{N}(A\mu + B, A\Sigma A^T)$