Announcements:

HW1 : submit via gradescope

Quiz1 : next Wednesday 9/27 (linear algebra, probability, optimization)

Waitlist and section changes : today & tomorrow

Outline

Recap. Const. Opt
Probability

$$p^* = \min_{x} f(x)$$
$$\text{subject to } g_i(x) \le 0 \quad \forall i \in \{1, \dots, m\}$$

Lagrangian:

$$\mathcal{L}(x, \lambda) = f(x) + \sum \lambda_i g_i(x) \qquad (\lambda_i \ge 0 \ \forall i)$$
$$= f(x) + \underline{\lambda}^T g(x)$$

Dual function

$$F(\lambda) = \inf_{x} \mathcal{L}(x, \lambda)$$

ex: $\inf_{x > 0} x = 0$ Infimum (inf): Greatest lower bound.

Dual problem: $d^* = \max_{\lambda \ge 0} F(\lambda)$

• $F(\lambda) \leq f(x^*)$ for any feasible x^* $\left(\min_{g_i(x) \leq 0} f(x) \right)$

$$F(\lambda) = \inf_x f(x) + \lambda^T g(x)$$
$$\leq f(x^*) + \lambda^T g(x^*)$$
$$\leq f(x^*)$$

$\Rightarrow$ $d^* \leq p^*$. (equal if nice enough)

Probability Terminology

Name	What it is	Common Symbols	What it means
Sample Space	Set	Ω, S	"Possible outcomes."
Event Space	Collection of subsets	$\mathcal{F}, E$	"The things that have probabilities.."
Probability Measure	Measure	P, π	Assigns probabilities to events.
Probability Space	A triple	$(\Omega, \mathcal{F}, P)$	

Remarks: may consider the event space to be the power set of the sample space (for a discrete sample space - more later). e.g., rolling a fair die:

$\Omega = \{1, 2, 3, 4, 5, 6\}$
$\mathcal{F} = 2^\Omega = \{\{1\}, \{2\} \ldots \{1, 2\} \ldots \{1, 2, 3\} \ldots \{1, 2, 3, 4, 5, 6\}, \{\}\}$

$P(\{1\}) = P(\{2\}) = \ldots = \frac{1}{6}$ (i.e., a fair die)
$P(\{1, 3, 5\}) = \frac{1}{2}$ (i.e., half chance of odd result)
$P(\{1, 2, 3, 4, 5, 6\}) = 1$ (i.e., result is "almost surely" one of the faces).

Axioms

$P(A) \geq 0 \quad \forall A \in \mathcal{F}$

$P(\Omega) = 1$

$P(\bigcup_i A_i) = \sum_{i=1}^n P(A_i)$ for disjoint events A_i $i \in \{1, \ldots, n\}$

Probability Mass Function (PMF) discrete R.V)

$f_X(x) \to P(X = x)$

Eg. $P(X = x_i) = \theta_i \geq 0$ $\quad \sum_{i=1}^{n} \theta_i = 1$

$\mathbb{E}\, g(X) = \sum_x g(x) f_X(x)$

Cumulative Distribution Function (CDF)

$F_X(x) = P(X \leq x)$

Probability Density Function (cont RVs)

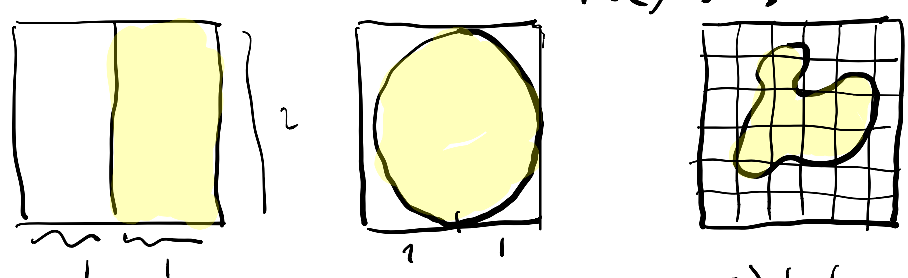
$f_X(x) = \frac{d}{dx} F_X(x)$ $\quad$ e.g. Uniform $f(x) = \frac{1}{b-a} 1_{(a,b)}(x)$

Gaussian

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

$N(\mu, \sigma^2)$

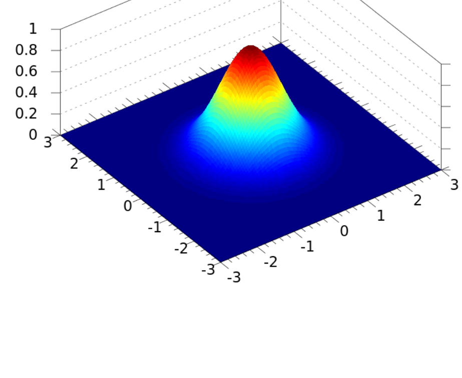
Uniforms:



Probability: $\frac{1}{2}$ $\quad \frac{\pi}{2 \cdot 2}$

a) Integration

b) Monte Carlo



Multivariate Gaussian

$\underline{X} = (X_1, \dots, X_n)$ is M.V.

Gaussian distributed if

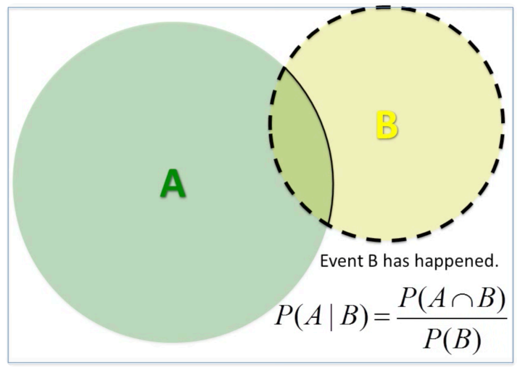
$$f_X(x_1,\dots,x_n) = \frac{1}{\sqrt{(2\pi)^n \cdot \det(\Sigma)}} \exp\left(-\frac{1}{2}\cdot (x-\mu)^T \Sigma^{-1} (x-\mu)\right)$$

$\underline{\mu} = \mu_1, \dots, \mu_n$ mean (vector)

$\underline{\underline{\Sigma}} \succeq 0$ (PSD) covariance (matrix)

Conditional Probability

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$



$$P(A|B) = \frac{P(B|A)\cdot P(A)}{P(B)}$$ Bayes Rule

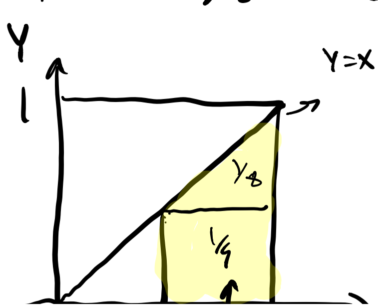
Independent R.Vs.

A and B are independent events if $P(A \cap B) = P(A)\cdot P(B)$

X and Y independent R.V.s if $P(X=x, Y=y) = P(X=x)\cdot P(Y=y)$

Guessing the bigger number: Observe X but not Y. in order to choose the bigger number will you pick X or Y?

X, Y uniform [0,1] independent



If $X > 1/2$ pick X
$X \leq 1/2$ pick Y

$X \leftrightarrow Y$

identical by symmetry

$$P(\text{win}) = P\left(x > \tfrac{1}{2} \cap x > y\right) + P\left(x < \tfrac{1}{2} \cap y > x\right)$$

$$= \frac{1}{4} + \frac{1}{8} + \frac{1}{4} + \frac{1}{8}$$

$$= 2 \cdot \frac{3}{8} = \frac{6}{8} = \frac{3}{4}$$

Classification

Consider random variables

$$X \in \mathbb{R}^d \quad (\text{feature vector})$$

$$Y \in \{1, \dots, K\} \quad \text{class label}$$

with joint distribution P_{XY}

$$P_{XY} = P_{X|Y} \cdot P_Y$$

$P_{X|Y}$: class conditional. P_Y: prior class distribution

$\pi_k = P_Y(Y = k) \quad k = 1, \dots, K \quad$ (discrete)

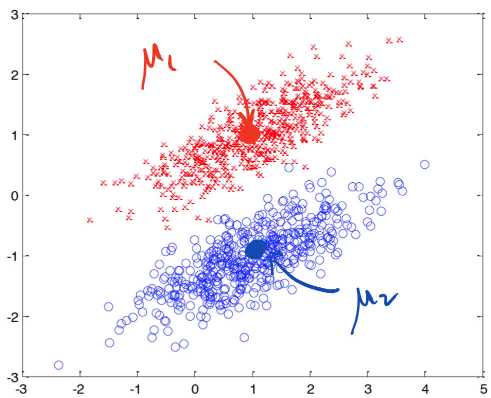
$P_{X|Y}$ discrete or continuous.

Ex: $K = 2$, $d = 2$ $\quad P_{X|Y}$ is Gaussian

$$X \mid Y = 1 \sim N(\mu_1, \Sigma)$$

$$X \mid Y = 2 \sim N(\mu_2, \Sigma)$$

$$\mu_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad \mu_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \quad \Sigma = \begin{bmatrix} .9 & .4 \\ .4 & .3 \end{bmatrix} \quad \pi_1 = \pi_2 = \tfrac{1}{2}$$



$$P_{XY} = P_{Y|X} \, P_X$$

↑ posterior distribution

- Naive Bayes: $P_{\underline{X}|Y=k}$ factorizes into independent factors.
- Linear Discriminant Analysis: $P_{X|Y=k}$ is Multivariate Gaussian.
- Logistic Regression: $P_{Y|X=x}$ is given by logistic probability model.

Bayes Classifier

Given P_{XY} of (X,Y) what is the best possible classifier

Classifier is a function

$$f: \mathbb{R}^d \to \{1, \dots, K\}$$

Performance measure: Probability of error (risk)

$$R(f) = P_{XY}(f(x) \neq y)$$