Lecture 1b. Probability Review

COMP90051 Statistical Machine Learning

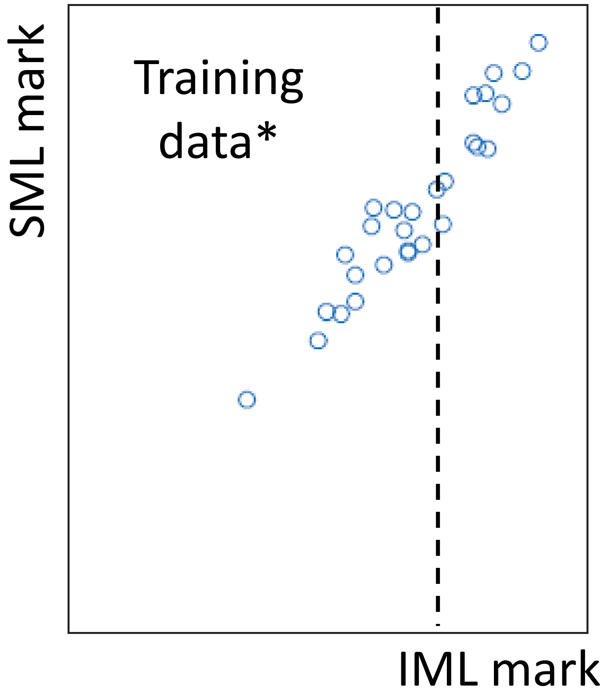
Sem2 2020
Lecturer: Ben Rubinstein

THE UNIVERSITY OF MELBOURNE

This lecture

- About COMP90051
- **Review: Probability theory**
- Review: Linear algebra
- Review: Sequences and limits

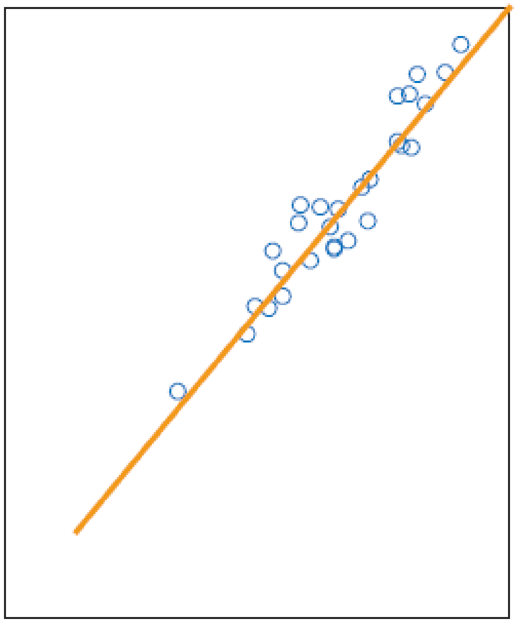
Data is noisy (almost always)



- Example:
  - given mark for Intro ML (IML)
  - predict mark for Stat Machine Learning (SML)

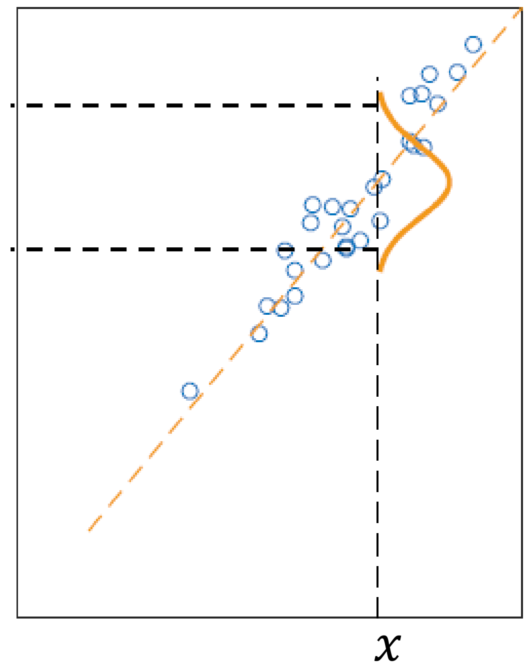
* synthetic data :)

Types of models



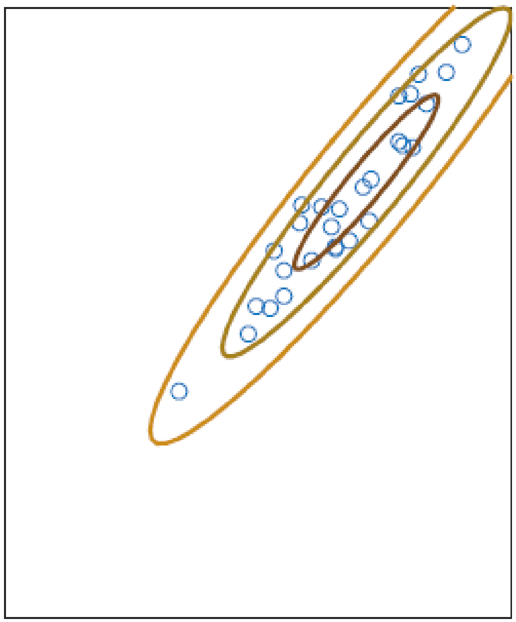
$\hat{y} = f(x)$

IntroML mark was 95, SML mark is predicted to be 95



$P(y|x)$

IntroML mark was 95, SML mark is likely to be in (92, 97)



$P(x, y)$

probability of having $(IML = x, SML = y)$

Basics of probability theory

- A probability space:
  - Set Ω of possible outcomes
  - Set *F* of events (subsets of outcomes)
  - Probability measure P: *F* → **R**
- Example: a die roll
  - {1, 2, 3, 4, 5, 6}
  - { φ, {1}, …, {6}, {1,2}, …, {5,6}, …, {1,2,3,4,5,6} }
  - P(φ)=0, P({1})=1/6, P({1,2})=1/3, …

Axioms of probability*

1. F contains all of: Ω; all complements $\Omega \backslash f$, $f \in F$; the union of any countable set of events in F.
2. $P(f) \geq 0$ for every event $f \in F$.
3. $P\left(\bigcup_f f\right) = \sum_f P(f)$ for all countable sets of pairwise disjoint events.
4. $P(\Omega) = 1$

* We won't delve further into advanced probability theory, which starts with measure theory – a beautiful subject and the only way to "fully" formulate probability.

Random variables (**r.v.'s**)

- A random variable X is a numeric function of outcome $X(\omega) \in \boldsymbol{R}$
- $P(X \in A)$ denotes the probability of the outcome being such that X falls in the range A

- Example: X winnings on \$5 bet on even die roll
  - X maps 1,3,5 to -5
    X maps 2,4,6 to 5
  - P(X=5) = P(X=-5) = ½

Discrete vs. continuous distributions

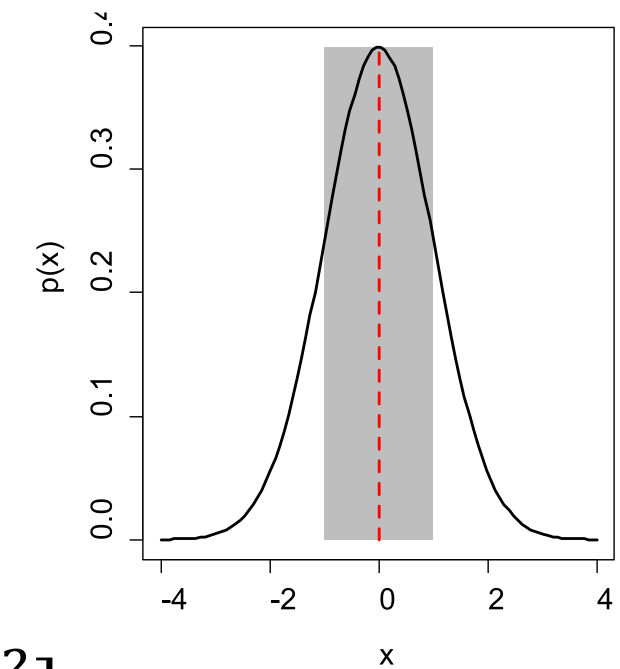
- Discrete distributions
  - Govern r.v. taking discrete values
  - Described by probability mass function p(x) which is P(X=x)
  - $P(X \leq x) = \sum_{a=-\infty}^{x} p(a)$
  - **Examples**: Bernoulli, Binomial, Multinomial, Poisson

- Continuous distributions
  - Govern real-valued r.v.
  - Cannot talk about PMF but rather probability density function p(x)
  - $P(X \leq x) = \int_{-\infty}^{x} p(a)da$
  - **Examples**: Uniform, Normal, Laplace, Gamma, Beta, Dirichlet

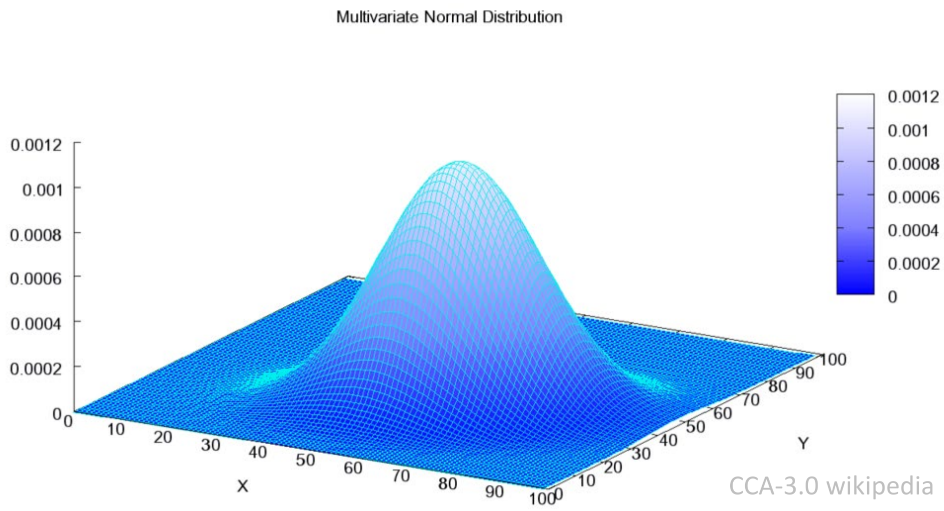
Expectation

- Expectation $E[X]$ is the r.v. X's "average" value
  - Discrete: $E[X] = \sum_x x\, P(X = x)$
  - Continuous: $E[X] = \int_x \, x\, p(x)\, dx$
- Properties
  - Linear: $E[aX + b] = aE[X] + b$

    $E[X + Y] = E[X] + E[Y]$
  - Monotone: $X \geq Y \;\Rightarrow\; E[X] \geq E[Y]$
- Variance: $Var(X) = E[(X - E[X])^2]$



Multivariate distributions



- Specify joint distribution over multiple variables
- Probabilities are computed as in univariate case, we now just have repeated summations or repeated integrals
- Discrete: $P(X, Y \in A) = \sum_{(x,y)\in A} p(x, y)$
- Continuous: $P(X, Y \in A) = \int_A \; p(x, y)dxdy$

Independence and conditioning

- *X*, *Y* are independent if
  - $P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$
  - Similarly for densities: $p_{X,Y}(x,y) = p_X(x)p_Y(y)$
  - **Intuitively**: knowing value of *Y* reveals nothing about *X*
  - **Algebraically**: the joint on *X*,*Y* factorises!

- Conditional probability
  - $P(A|B) = \frac{P(A \cap B)}{P(B)}$
  - Similarly for densities $p(y|x) = \frac{p(x,y)}{p(x)}$
  - **Intuitively**: probability event *A* will occur given we know event *B* has occurred
  - X,Y independent equiv to $P(Y = y|X = x) = P(Y = y)$

Inverting conditioning: Bayes' Theorem

Bayes

- In terms of events *A*, *B*
  - $P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$
  - $P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$
- Simple rule that lets us swap conditioning order
- Probabilistic and Bayesian inference make heavy use
  - Marginals: probabilities of individual variables
  - Marginalisation: summing away all but r.v.'s of interest

    $P(A) = \sum_b P(A, B = b)$

Summary

- Probability spaces, axioms of probability
- Discrete vs continuous; Univariate vs multivariate
- Expectation, Variance
- Independence and conditioning
- Bayes rule and marginalisation

Next: Linear algebra primer/review