Subject Review and Exam Info

COMP90051 Statistical Machine Learning

Lecturer: Feng Liu

THE UNIVERSITY OF MELBOURNE

In the final lecture

- We will quickly review main contents of this subject
- Exam paper info
  - Exam scope and exam info
  - Changes in this semester, some points need your feedback to improve SML in future
  - Resources for exam prep
- Please consider*!!* End of Semester Surveys (ESS)

Subject Review

What you have learnt in this subject

Main contents in SML (Basis)

lec 1 - 5 basic concepts in ml

- Stats background and thoughts – fundamental parts -- lectures 1 and 2 frequentist and bayesian
- Linear regression -- lecture 3
  - Simple model (convenient maths at expense of flexibility)
  - Often needs less data, "interpretable", lifts to non-linear
- Basis expansion: Data transform for more expressive models -- lecture 3

Main contents in SML (Basis)

non-linear part in logistic regression

- Logistic regression: workhorse linear classifier -- lecture 4
  - Possibly familiar derivation: frequentist
  - Decision-theoretic derivation
  - Training with Newton-Raphson looks like repeated, weighted linear regression
- Regularising linear regression -- lecture 5
  - Ridge regression
  - The lasso
  - Connections to Bayesian MAP
- Regularising non-linear regression -- lecture 5
- Bias-variance -- lecture 5

Main contents in SML (ML Theory) Lecture 6

function class:
very large -> best model might be within this class
big function class: estimation error high, approx. error low
small function class: estimation error low, approx. error higher (
easy to find the best function within this class)

- Excess risk
  - Decomposition: Estimation vs approximation
  - Bayes risk – irreducible error

PAC learning theory, important!

- Probably approximation correct learning
- Bounding generalisation error with high probability
  - Single model: Hoeffding's inequality
  - Finite model class: Also use the union bound

    finite functions within the class
- Importance & limitations of uniform deviation bounds

Bound for finite classes $\mathcal{F}$

- A uniform deviation bound over *any* finite class or distribution

Theorem: Consider any $\delta > 0$ and finite class $\mathcal{F}$. Then w.h.p at least $1 - \delta$: For all $f \in \mathcal{F}$, $R[f] \leq \hat{R}[f] + \sqrt{\frac{\log|\mathcal{F}| + \log(1/\delta)}{2m}}$

true risk

empirical risk (smth we can calculated thrugh the sample)

Proof:

- If each model f having large risk deviation is a "bad event", we bound the probability that any bad event happens.
- $\Pr\left(\exists f \in \mathcal{F}, R[f] - \hat{R}[f] \geq \varepsilon\right) \leq \sum_{f \in \mathcal{F}} \Pr\left(R[f] - \hat{R}[f] \geq \varepsilon\right)$
- $\leq |\mathcal{F}| \exp(-2m\varepsilon^2)$ by the union bound
- Followed by inversion, setting $\delta = |\mathcal{F}| \exp(-2m\varepsilon^2)$

Main contents in SML (ML Theory) Lecture 7

- PAC learning bounds:
  - Countably infinite case works as we've done so far
  - General infinite case? Needs new ideas!
- Growth functions for the general PAC case
  - Considering patterns of labels possible on a data set
  - Gives good PAC bounds provided possible patterns don't grow too fast in the data set size
- Vapnik-Chervonenkis (VC) dimension we can measure the size of infinite function class
  - Max number of points that can be labelled in all ways
  - Beyond this point, growth function is polynomial in data set size
  - Leads to famous, general PAC bound from VC theory
- Optional proofs at end (just for fun)

PAC Bounds

- **Theorem** (finite class): Consider any $\delta > 0$ and finite class $\mathcal{F}$. Then w.h.p at least $1-\delta$: For all $f \in \mathcal{F}$, $R[f] \leq \hat{R}[f] + \sqrt{\frac{\log|\mathcal{F}| + \log(1/\delta)}{2m}}$

why we can rely on ml models based on empirical result:

- Theorem (VC bound): Consider any $\delta > 0$ and any VC-k class $\mathcal{F}$. Then w.h.p. at least $1-\delta$: For all $f \in \mathcal{F}$

$$R[f] \leq \hat{R}[f] + 2\sqrt{2\frac{k\log\frac{2em}{k} + \log\frac{4}{\delta}}{m}}$$

Main contents in SML (SVM)

a classical learning framework in ml

Lecture 8 and Lecture 9

- Support vector machines (SVMs) as maximum-margin classifiers – lecture 8
- The hard/soft-margin SVM objective -- lecture 8 what we want to do
- Dual formulation of the SVM -- lecture 9 how to solve svm
- Kernelisation (solution to SVM) -- lecture 9
  - Basis expansion on dual formulation of SVMs
  - "Kernel trick"; Fast computation of feature space dot product
- Constructing kernels -- lecture 9
  - Overview of popular kernels and their properties
  - Mercer's theorem
  - Learning on unconventional data types

Main contents in SML (Neural Network) Lecture 10,11,12,13,14,15

- Perceptron model and its training procedures (basis of NN) – lecture 10
- Fundamentals -- lecture 11
  - Networks, layers, activation functions
  - Training by gradient backpropagation
- Training & Autoencoders -- lecture 12
- Network architectures 6 lectures
  - Convolutional networks (CNN) -- lecture 13
  - Recurrent networks (RNNs) -- lecture 14 briefly shown
  - Attention and the Transformer -- lecture 14
  - Graph NN -- lecture 15

Main contents in SML (RL) Lecture 16 and Lecture 17

- Learning from expert advice / multiplicative weights
- Infallible expert (one always perfect)
- Imperfect experts (none guaranteed perfect)
- Bandit setting vs Learning with experts
- Aka. Sequential decision making under uncertainty
- Basic algorithms (Greedy, ε-Greedy, Upper Confidence Bound (UCB))

Main contents in SML (Bayesian ML) Lecture 18 and Lecture 19

- Bayesian regression – lecture 18 briefly introduced
  - Sequential Bayesian updating
  - Conjugate prior (Normal-Normal) prior and likelihood are normal -> posterior is normal as well
  - Using posterior for Bayesian predictions on test

A prior is called conjugate to a likelihood if the resulting posterior distribution is of the same family as the prior.

- Bayesian classification – lecture 19
  - Beta-Binomial conjugacy
  - Uniqueness up to proportionality ??
  - Rejection sampling -- Monte Carlo sampling

how we get sample from un-normalised distribution
we reject some sample from the uniform distribution

Main contents in SML (PGMs) Lecture 20,21

graph models

- Direct PGMs – lecture 20
  - Independence lowers computational/model complexity
    we can reduce the size of table by joint distribution based on indept.
    - Conditional independence
  - PGMs: compact representation of factorised joints
    complexity
    k r.v., non-indept: 2^k - 1
    k r.v. indept: k
- Undirect PGMs – lecture 21
  - Undirected PGM formulation
  - Directed to undirected

Main contents in SML (PGMs) Lecture 22

- Probabilistic inference: computing (conditional) marginals from joint distributions
  - Needed to learn (posterior update) in Bayesian ML
    naive way will cost computation resources
  - Exact inference: Elimination algorithm
  - Approximate inference: Sampling
    based on joint distribution: we can do some quick inference!
- Statistical inference: Parameter estimation
  - Fully observed case: Factors decompose under MLE
  - Latent variables: Motivates the EM algorithm

Main contents in SML (PGMs) Lecture 23

- Gaussian mixture model (GMM) special case of direct PGM
  - A probabilistic approach to clustering how we model clustering problem in PGM
  - The GMM model
  - GMM clustering as an optimisation problem
- Briefing Expectation-Maximisation (EM) algorithm

Main contents in SML

- Seven parts:
  - Basis – lectures 1-5
  - Machine Learning Theory -- lectures 6-7
  - SVM&Kernel -- lectures 8-9
  - Neural Networks -- lectures 10-15
  - Reinforcement Learning -- lectures 16-17
  - Bayesian Machine Learning -- lectures 18-19
  - PGMs (including clustering) -- lectures 20-23

Exam Info

Exam scope

Resources for exam prep

Exam Scope in SML (Basis)

- Stats background and thoughts – fundamental parts -- lectures 1 and 2)
- Linear regression -- lecture 3
  * MLE is important
- Regularising linear regression -- lecture 5
  * Ridge regression
  * The lasso
  * Connections to Bayesian MAP
- Regularising non-linear regression -- lecture 5
- Bias-variance -- lecture 5

Exam Scope in SML (ML Theory) Lecture 7

- PAC learning bounds:
  - Countably infinite case works as we've done so far
  - General infinite case? Needs new ideas!
- Growth functions for the general PAC case
  - Considering patterns of labels possible on a data set
  - Gives good PAC bounds provided possible patterns don't grow too fast in the data set size
- **Vapnik-Chervonenkis (VC) dimension** VC dimension!!! very important
  - Max number of points that can be labelled in all ways
  - Beyond this point, growth function is polynomial in data set size
  - Leads to famous, general PAC bound from VC theory

Exam Scope in SML (SVM) Lecture 9

- **Kernelisation (solution to SVM)**
  - Basis expansion on dual formulation of SVMs
  - “Kernel trick”; Fast computation of feature space dot product

    inner product of 2 functions

    how to find the mapping function

- Constructing kernels
  - Overview of popular kernels and their properties
  - Mercer’s theorem
  - Learning on unconventional data types

Exam Scope in SML (Neural Network) Lecture 11,12,13,14

- Fundamentals -- lecture 11
  - Networks, layers, activation functions
  - Training by gradient backpropagation

graph NN will not be included

- Training & Autoencoders -- lecture 12
- Network architectures
  - **Convolutional networks (CNN) -- lecture 13**
  - Recurrent networks (RNNs) -- lecture 14

Exam Scope in SML (RL) Lecture 16 and Lecture 17

- Not included

Lecture 16. Learning with expert advice,
Lecture 17. Multi-armed bandits

Exam Scope in SML (Bayesian ML) Lecture 18 and Lecture 19

- Not included but Bayes theorem is important!!

$$P(\theta|X) = \frac{P(\theta, X)}{P(X)}$$

Exam Scope in SML (**PGMs**)
Lecture 20,21 !!!

- Direct PGMs – Lecture 20
  - Independence lowers computational/model complexity
    - Conditional independence   elimination algorithm
  - PGMs: compact representation of factorised joints
- Undirect PGMs – Lecture 21
  - Undirected PGM formulation
  - Directed to undirected

Exam Scope in SML (PGMs) Lecture 22

if you have direct PGM, how to eliminate the random variable one by one to make the quicker prob inference

- Probabilistic inference: computing (conditional) marginals from joint distributions
  - Needed to learn (posterior update) in Bayesian ML
  - Exact inference: Elimination algorithm
  - Approximate inference: Sampling

show direct PGM, we can only observe rv1, rv2 how to make statistical inference?

- Statistical inference: Parameter estimation
  - Fully observed case: Factors decompose under MLE
  - Latent variables: Motivates the EM algorithm

lec23 will not be included

Exam Scope in Application

1. short question
2. method
3. application

- Project 1 is important – review your own project and think about the procedures when you handle this project.

Top Tips during the Exam

- Don't panic! easier than last year
- Attempt all questions
  - Do your best guess whenever you don't know the answer
- Finish easy questions first (do q's in any order)
- If you can't answer part of the question, skip over this and do the rest of the question
  - you can still get marks for later parts of the question
  - we don't repeatedly penalise for carrying errors forward

"Open book" format

- What you can and can't use:
  - can = lecture materials, reading, workshops, your project reports, previous exams, electronic calculators
  - can't = web resources, web search, electronic messaging, collaboration with others, private tutor notes, ChatGPT
- But… set similar like prior years'
  - there's not a lot of time to be reading through materials
  - don't memorise the mathematical formula
    (can use 2019 exam formula page on Canvas if desired)
  - few definitional questions, more conceptual & worked problems

Changes in this semester

- Contents

  hidden markov model

  - Talk less about HMM

  + Rejection sampling

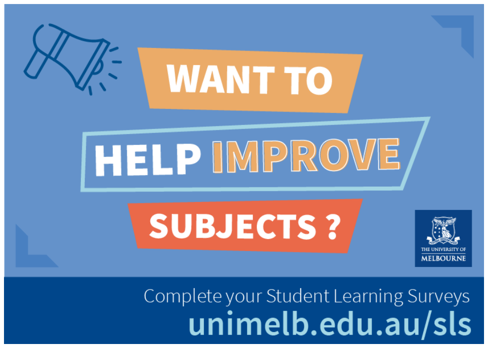
  basic and important

  + More focus on deep understanding instead of wide contents, try step-by-step derivations

We do need your comments on these changes!!

- Assignments
  - Detailed feedback + feedback session during lecture
  - Advanced Project 1 (play on your own with the released datasets)
  - Reduce the scale of Assignment 2

Need your comments via ESS



We do need your comments on these changes!!

Way 1 (SLS): https://www.unimelb.edu.au/sls/

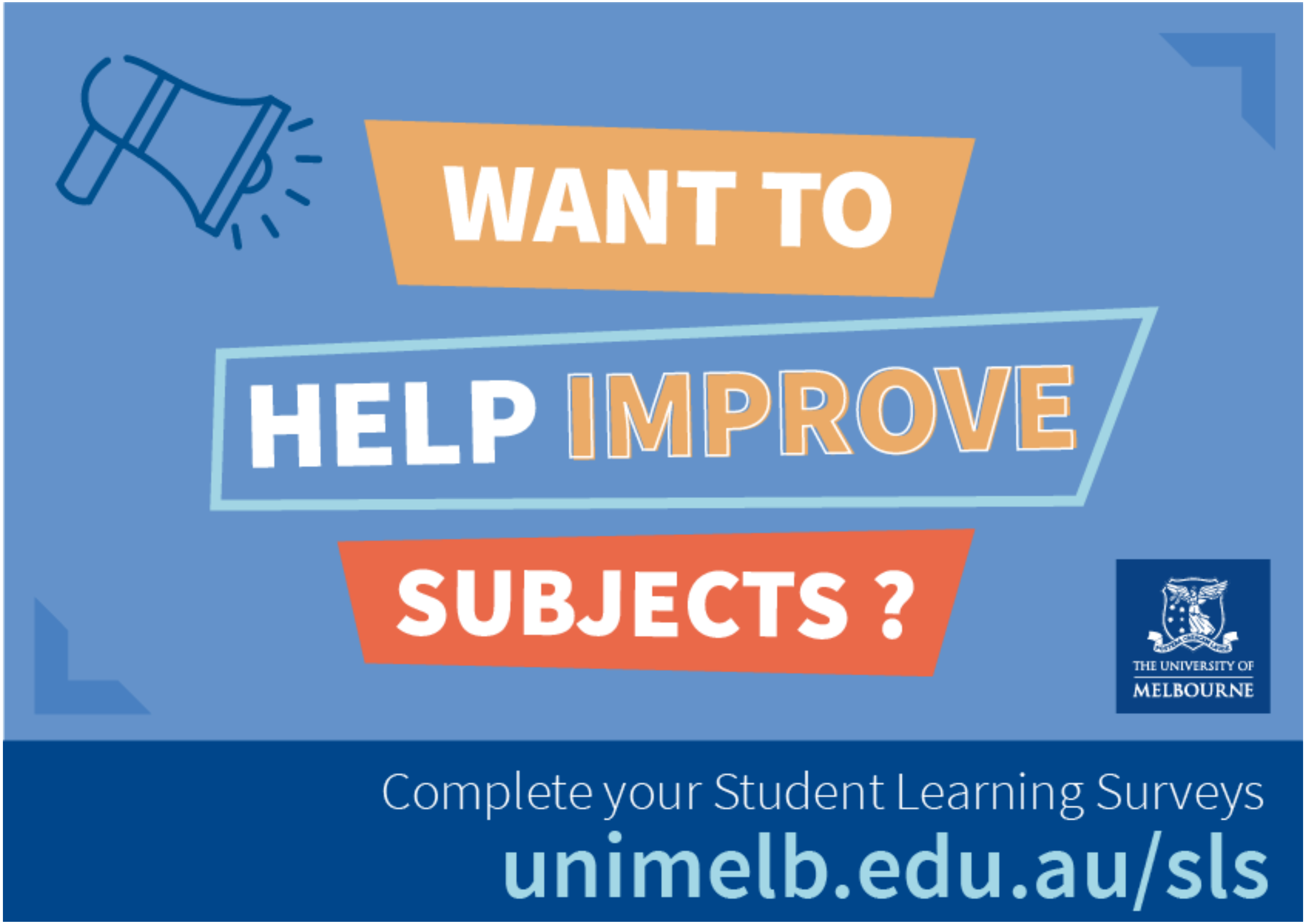
Way 2 (LMS): https://www.unimelb.edu.au/sls/students/ess

Please do complete the **End of Semester Survey (ESS)**. It really only takes you a few minutes, but means ***a lot*** to us - whether you have ***positive*** or ***negative*** or ***mixed*** feedback. Please don't assume that other students will respond for you, we really need everyone to respond to get a complete picture of how the subject is going and if the changes are helpful. Thanks!

Resources for exam prep

- Practice exam – for format
- Past exams in library – calibrate for difficulty
  - We don't provide solutions for past exams, sorry
- Quizzes
- Workshops, lectures, readings
- Ed to clarify concepts from now until exam
- Two face-to-face exam consultation sessions (details will be included in the next announcement)

Good luck, you'll do great!


WANT TO
HELP IMPROVE
SUBJECTS ?
THE UNIVERSITY OF
MELBOURNE
Complete your Student Learning Surveys
unimelb.edu.au/sls