Exam Preperration Session

Lecture 1:

- important: multivariate guassian
- know difference normal & gradient descent
  - question: if I swap those parameters can I still use?
  - when can you not use normal equation. If normal equation so good, why do we still look at Gradient Descent.
- regularization (not very important but know)
- softmax regression and how you use it!
  - when k-binary classifier, whats the difference
- recommender systems (!)
  - content based approach know whats going on /algorithm and know the computations.
  - kurt will ask you something out of the box, e.g. apply the algorithm to classify. you really need to know algorithm well.
- NNs
  - explainability

Lecture 2: Boltzmann + Deep Learning

- KL-divergence and cost function
- Variational Autoencoders → Latent Space (!)
- know the algorithm
- vanishing gradient problem
  - what happens? what causes this? know it very well
  - what is the solution to this problem
- RNN is not that important, but know when to use this.
- check self-attention
- hopfield net not really in exam
- how to calc boltzmann, the energy is important, how you calculate the probabilities. Why is derivative simple?
- kurt is fan of hinton. Positive & Negative phase, understand this deeply. Slide 28 & 29

- log reg kernal
- ~~should we read the books~~
- passing rate

Lecture 3: SVM:
- wont ask to derive mathematics
- what if dataset is an image, how would you apply SVM. Know the kernal trick
- slack variable, what does it control?
- kernel trick past exams very important
- SMO probably not asked, understand concept

Lecture 4: Gaussian Process
- watch the videos about it!
- why is good. How do we construct method?
- how do we derive is
- μ^*, σ^* MOST IMPORTANT   know complexity

Lecture 5: Representation
- always first question on exam
- know difference between methods
- always explain WHY you use this method and not others
- boolean, AV, MI, RR rest less important
- kernels applied to multi instance

Lecture 7:
- design a specific kernel so you can solve this problem.
- no questions probably about graph kernels
- clustering medium chance of appearing
  - ↳ set matching neither

Lecture 8: Quantum ML
- know the gates (mostly for assignments)
- slide 56: overview physical circuit
  - ↳ could be asked to draw something like this
  - ↳ know what each part does and means and how they work
  - ↳ how does data pass between them
- Variational Quantum Classifier
- Feature maps
- Comparison quantum NN & normal NN. (!!!)

Previous Exam 2020-2021

Q1: Representational Issues

We want to save ~~the~~ collisions in CERN with properties & and their label interesting / non-interesting.

ASSUMPTIONS: e.g. ~~why~~ how traj. snapshot is represented

I will use Multi-Instance to represent this problem:

		Particle	momentum	charge (v)	traj. snapshot	Label
i.e.	p1	μ	4	3	$\begin{bmatrix} 0 \\ 3 \end{bmatrix}$	interesting
	p2	σ	5	1	$\begin{bmatrix} 4 \\ 2 \end{bmatrix}$	
	p3	θ	1050	0.5	[ ]	
	p4	φ	1	0.5		
i.e.	p5	ψ	50	0.01		not interesting
	p6	μ	0.01	50		

a learning example represents a collision here, where the rows represent 1 particle per collision.

BY ⊂ AV ⊂ MI ⊂ RR

Why use Multi-Instance? In multi-instance I can represent multiple objects attached to one learning example. This means I can save a row per particle created in a collision. If we move one representation level down, to Attribute value we lose the ability

We are not using RR cause we gain information but not necessary + less solvers + less complex better

Q2: CAE'S

a) Explain why/how a shallower network can perform better in this learning setting.

A shallower network means there is less convolution/averaging, thus leaving the latent space bigger, & so we have more space to represent our data, thus this can perform better.

b) would you use this in the colorization task

No, as you want it to generalise over features. For colorization we don't want specific details of images. I ar

c) how can we increase the reconstruction/colorization

more channels, smaller kernels.

Q3: Kernel Design

So our main issue is categorical data (?) so we can represent male female like 0,1 and the hair colors on a numerical scale with 0 black and max white and scale on darkness.

we need to define a distance metric

1. transfer to numerical
2. transpose to 4xy matrices as we have 4 features. So it against its transpose

Q4: Softmax & SSD

a) Do you think using an SSD loss for an NN with softmax final layer is possible?

Softmax outputs a probability for each class. SSD uses the selected class. Without adaptations, this is not possible. So we compare a probability and a ground truth. We could argmax our softmax but then its not
b) ~~Ye we can change~~ really softmax anymore.

c) Yes, you could but I'm not sure at all tbh you measure the probability error. You could, but its weird... But maybe not.