Machine Learning 1

Sample Questions 2

WS18/19 (inofficial Sample questions for Friday, February 1st)

Name:

Subject of study:

Tutorial you attended:

☐ Alexander Lenz / Kashef Karim

☐ Tristan Scheidemann

Points:

Q.1	Q.2	Q.3	Q.4	Q.5	Q.6	Q.7	Q.8	Q.9	$\sum$

Q.1 (10 points)

Please answer each of these questions briefly in full sentences. Each correctly answered question adds 2 points to the score. Wrong answers do not subtract points.

Name 3 different gradient methods and explain briefly (in words) what they do.

In linear regression, we noticed that the Bayesian posterior $p(\mathbf{w}|\mathbf{X}, \mathbf{t})$ approaches the maximum likelihood solution if the number of samples goes to infinity. Briefly explain why this is the case.

Briefly explain why the Heaviside step function is generally not used for deeper Neural Networks that are trained through backpropagation.

Let us do the question from last time again! Imagine you have n samples and a parameter θ. You now receive m additional datapoints. Explain the role of "Bayesian updating" in context of the new posterior $p(\theta|\mathbf{X}_{new}, \mathbf{X}_{old}) \propto \cdots$.

Selection of a good prior is essential for good estimates if we have little data. In the lecture, we learned different paradigms for the selection of prior distributions. Briefly explain what they mean:

- Selection based on conjugacy (conjugate prior)
- Selection based on normalization (proper vs improper priors)
- Selection based on invariance principles (e.g. Jeffrey's prior)

Q.2 (10 points)

We are throwing n independent coins with probability $\theta = p$. We are interested in finding θ.
Last time we deduced the likelihood:

$$L(\theta) = \binom{n}{k} \theta^{k} (1 - \theta)^{n-k}$$

and maximum likelihood solution

$$\theta_{MAX} = \frac{k}{n}, \qquad where\ k = \#heads$$

We are now interested in finding the solution for the Bayesian formulation:

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

Q.2.1 (5 points)

Assume a $Beta(3,7)$ prior. How is the posterior distributed?

Hint:

$$Beta(\theta; a, b) = c \cdot \theta^{\alpha-1}(1-\theta)^{\beta-1}, \qquad c \in \mathbb{R}^{+}.$$

Q.2.2 (5 points)

We receive the following datapoints:

$$X = (1,1,1,0,0,1,0,1)$$

Compare the expected value of θ for the maximum likelihood solution and Bayesian formulation given the above data.

Hint:

$$E[X] = \frac{\alpha}{\alpha + \beta}, \qquad if\ X \sim Beta(a, b)$$

Q.3 (5 points)

Design an arbitrary neural network f with two inputs X_1, X_2 that realizes an AND gate. Your activation functions must be Heaviside, e.g.

$$h(a)_i^l = \begin{cases} 1, & if\ a \geq 0 \\ 0, & else \end{cases}$$

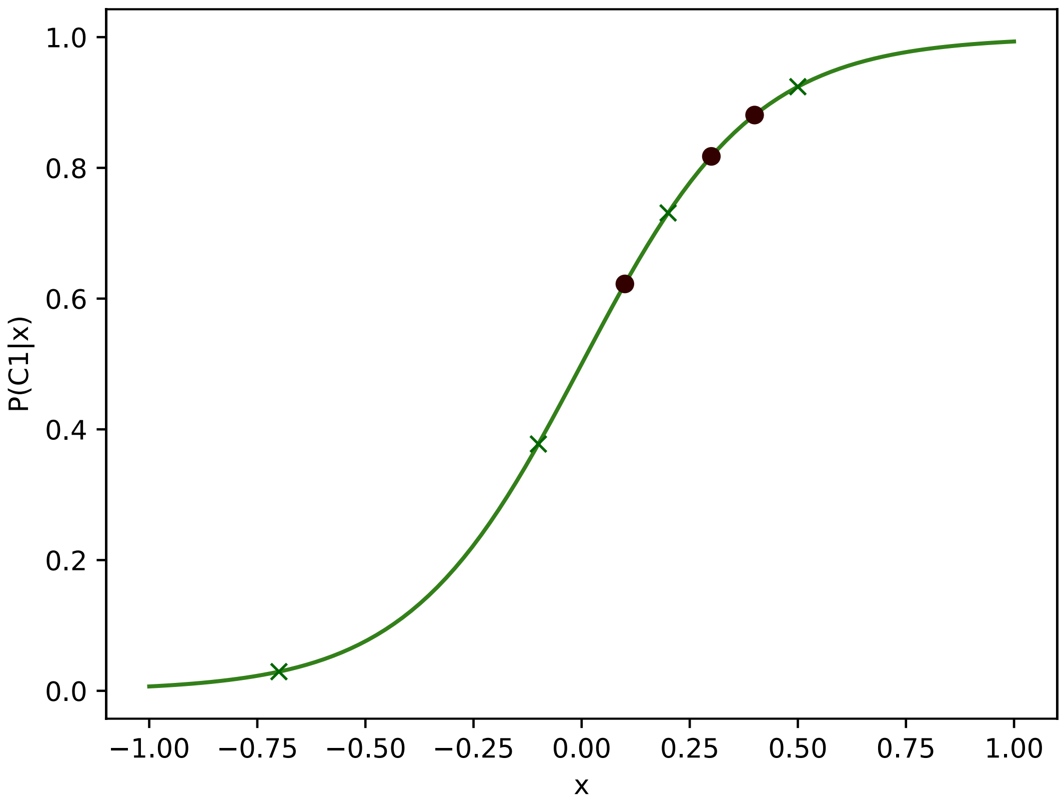
The AND gate has the following truth table.

X_1	X_2	$X_1 \wedge X_2$
0	0	0
0	1	0
1	0	0
1	1	1

Hint:
A single layer network is enough.

Q.4 (20 points)

Given is the following binary data in a classification problem. The circles belong to class 1 and the crosses to class 2.



Drawn is $P(C_1|x) = \frac{1}{1+e^{-5x}}$. We classify x belonging to C_1 if it satisfies $P(C_1|x) > \theta$.We want to see how the quality of estimates changes when we differ θ.

In other words: Sketch the ROC curve.