Vertiefungslinie Intelligent Systems SS21

Machine Learning and Reinforcement Learning

Examiners: Steffen Staab (ML) and Jim Mainprice (RL)
Exam Duration: 45 min
Date: 29.09.2021

General Remarks

- This is written from memory and might be incomplete and inaccurate.
- The exam was an oral exam. The examiners asked questions which I answered verbally or on a blackboard with chalk.
- For ML I got to choose between English or German. RL had to be answered in English.
- After the exam I was asked to briefly go out of the room while they decided my grade.
- I got the grade 2,0.
- Solutions are at the end of this document.

Machine Learning

How can you specify the classification problem in terms of probability?

How do we get from towards Naïve Bayes (first without naïve assumption)?

Which assumption do we make in Naïve Bayes?

Why do we use this assumption? Why not use the correct Formula?

Which Problem do we have with Naïve Bays? How do we solve is?

What is special about Naïve Bayes text classification?

We inherently have an irreducible classification error. Why?

Which methods did we learn to perform a regression?

What is an artificial neural network?

Give an equation of an Artificial Neural Network if we want to use it for regression.

Why do we need the activation functions?

What are examples for an activation function?

What is a loss function?

What is an example for a loss function that we can use?

Why does MSE work good for regression?

Which properties should a loss function have?

How can we train a NN?

How do we calculate the gradient?

What is overfitting?

How can we prevent overfitting? How does it work?

Reinforcement Learning

What is Reinforcement Learning?

What is a Markov Decision Process? Write the components of an MDP.

What is a v-function? Give the definition.

How is this equation called?

What is special about it? (First thing somebody would notice)

What is q-function? Give the definition.

Write the update step of value iteration.

We learnt Dynamic Programming approaches. Are they considered Reinforcement Learning?

Write the update rule of SARSA. Which part of this is the TD-Target?

Is SARSA on-policy or off-policy? Why?

Is Q-Learning on-policy or off policy? Why?

Which kind of policy do we typically use as a behavior policy? Why?

What is function approximation in RL?

If we use a linear function for function approximation of v. What would be the update rule?

How do policy gradient methods work?

Solutions Machine Learning

How can you specify the classification problem in terms of probability?

$$\mathrm{f(o)} = \underset{c}{\operatorname{argmax}} P(c|o)$$

How do we get from towards Naïve Bayes (first without naïve assumption)?

$$\begin{aligned} P(c|o) &= \frac{P(c) \cdot P(o|c)}{P(o)} \\ &= \frac{P(x_1, \dots, x_m|c) \cdot P(c)}{P(x_1, \dots, x_m)} \\ &= \frac{P(x_1, \dots, x_{m-1}|x_m, c) \cdot P(x_m|c) \cdot P(c)}{P(x_1, \dots, x_m)} \end{aligned}$$

Which assumption do we make in Naïve Bayes?

We assume that the label only depends bi-laterally on the attribute values.

$$P(c|o) = \frac{P(x_1, \dots, x_m|c) \cdot P(c)}{P(x_1, \dots, x_m)} = \frac{P(x_1|c) \cdot \dots \cdot P(x_m|c) \cdot P(c)}{P(x_1, \dots, x_m)}$$

Why do we use this assumption? Why not use the correct Formula?

Measuring probabilities $P(x_1, \dots, x_{m-1}|x_m, c)$, $P(x_1, \dots, x_{m-2}|x_{m-1}, x_m, c)$, … becomes impossible as not all cases will be available in training set.

Which Problem do we have with Naïve Bays? How do we solve is?

If only one $P(x_i|c)$ is zero, then $P(c|x_1, \dots, x_m)$ will be zero too.

This can be solved by smoothing (Laplace, Lidstone).

What is special about Naïve Bayes text classification?

We only consider terms that are in a document, but not terms that aren't. Terms which are not present do not affect the class association.

We inherently have an irreducible classification error. Why?

Due to variance of attributes $P(c|o)$ is seldom 0 or 1. And correct class might be the non-maximum. For example, men are usually taller than women, but there are also tall women and short men. If someone is 182 cm tall, it's most likely a man but also possibly a woman.

Which methods did we learn to perform a regression?

Linear regression, Neural Network.

What is an artificial neural network?

Combination of linear functions with non-linear activation functions

Give an equation of an Artificial Neural Network if we want to use it for regression.

$f(x) = W_n(\dots a(W_2 a(W_1 x + b_1) + b_2) \dots) + b_n$

Why do we need the activation functions?

To introduce non-linearity, because otherwise combining to linear functions gives us another linear function. $W_2(W_1 x) = (W_2 W_1)x$

What are examples for an activation function?

Sigmoid function $\sigma(x) = \frac{e^x}{1+e^x}$, ReLU

What is a loss function?

A measure of how bad a classification or regression is.

What is an example for a loss function that we can use?

For regression we can use Mean Squared Error

Why does MSE work good for regression?

Loss is small if prediction is close to training data.

Which properties should a loss function have?

Differentiable, Convex to be able to perform gradient descent.

How can we train a NN?

Using Backpropagation.

How do we calculate the gradient?

We calculate $\frac{\delta L(x)}{\delta W}$ by calculating the gradients step by step: $\frac{\delta L(x)}{\delta W_1} = \frac{\delta L(x)}{\delta y} \cdot \frac{\delta y}{\delta W_1}$

What is overfitting?

The model fits too much to the training data, that it generalizes badly for unseen data.

How can we prevent overfitting? How does it work?

Regularization. Introducing bias to reduce variance.

Solutions Reinforcement Learning

What is Reinforcement Learning?

RL is an area of machine learning where the goal is to train an agent's behavior using samples taken from an environment (typically Markov Decision Process) to maximize a reward.

What is a Markov Decision Process? Write the components of an MDP.

State $s \in S$, action $a \in A$, reward $r \in R$

Model $p(s', r|s, a)$, Policy $\pi(a|s)$

What is a v-function? Give the definition.

The v-function gives for each state the expected (discounted) future reward under policy π.

$v_\pi(s) = \mathbb{E}_\pi[G_t|S_t = s] = \sum_a \pi(a|s) \sum_{s',r} p(s', r|s, a)[r + \gamma v_\pi(s')]$ for all $s \in S$

How is this equation called?

Bellman equation

What is special about it? (First thing somebody would notice)

It's recursive.

What is q-function? Give the definition.

The q-function gives for each state-action-pair the expected (discounted) future reward under policy π.

$q_\pi(s) = \mathbb{E}_\pi[G_t|S_t = s, A_t = a] = \sum_{s',r} p(s', r|s, a)[r + \gamma \sum_{a'} \pi(a'|s') q_\pi(s', a')]$ for all $s \in S$

Write the update step of value iteration.

$v_\pi(s) = \max_a \sum_{s',r} p(s', r|s, a)[r + \gamma v_\pi(s')]$

We learnt Dynamic Programming approaches. Are they considered Reinforcement Learning?

It's not considered reinforcement learning. (I didn't know this)

Write the update rule of SARSA. Which part of this is the TD-Target?

$$Q(S, A) \leftarrow Q(S, A) + \alpha \left[\underbrace{R + \gamma Q(S', A')}_{\text{TD-Target}} - Q(S, A) \right]$$

Is SARSA on-policy or off-policy? Why?

On-policy because the target policy is the behavior policy.

Is Q-Learning on-policy or off policy? Why?

Off-policy because the target policy is different from the behavior policy.

Which kind of policy do we typically use as a behavior policy? Why?

An ϵ-greedy policy to allow for exploration and potentially find a better policy.

What is function approximation in RL?

If we want to have big or even continuous state or action spaces then we can't store the q-, v- and π-functions in a table. Instead, we can try to approximate the functions. For example, using linear functions or neural networks.

If we use a linear function for function approximation of v. What would be the update rule?

$w_{t+1} = w_t + \alpha[U_t - w_t^T x(S_t)] \cdot x(S_t)$ with U_t being the TD-Target

How do policy gradient methods work?

In policy gradient methods we try to find the optimal policy by performing gradient ascent on the policy. (I didn't know the answer and I'm not sure what best answer would be)