

2020/21, Primeiro teste de AA / AA first test.

Pergunta 1

Português

Atenção: a sua resposta deve ser concisa e bem estruturada. Se vir que não tem tempo para dar uma resposta estruturada que cubra tudo o que é pedido, foque apenas os itens mais importantes.

Suponha que tem um conjunto de dados com 1000 pontos do tipo (x, y) e quer ajustar um modelo de regressão para prever o valor contínuo de y para novos exemplos com base nos valores de x . Quer seleccionar o melhor de vários modelos e precisa de uma estimativa do erro que em média espera ter quando usar a sua hipótese final em exemplos novos. Explique como procederia para resolver este problema. Deve focar os seguintes aspectos, por ordem decrescente de prioridade:

- Como organiza este conjunto de dados em diferentes conjuntos, explicando porquê.
- Como selecciona o melhor modelo.
- Como obtém a hipótese final.
- Como obtém uma estimativa não enviesada do erro de previsão para exemplos novos.
- Qual a função de erro que vai usar, justificando a escolha.

English:

Attention: your answer should be concise and well structured. If you feel you do not have time to give a well structured answer covering all items below focus on the most important ones.

Suppose you have a data set of 1000 points of the form (x, y) and want to fit a regression model to predict the continuous value y for new examples based on the values of x . You want to select the best of several models and you need an estimate of the average expected error when using your final hypothesis on new examples. Explain how you would proceed to solve this problem. You should focus on the following aspects, in decreasing order of importance:

- How you organize this data set in different sets, explaining why.
- How you select the best model.
- How you obtain the final hypothesis.
- How do you get an unbiased estimate of the expected error for new examples..
- What is the loss function you will use, and justify this choice of loss function.

Pergunta 2

Português

Atenção: a sua resposta deve ser concisa e bem estruturada. Se vir que não tem tempo para dar uma resposta estruturada que cubra tudo o que é pedido, foque apenas os itens mais importantes.

A Regressão Logística é um classificador linear mas pode ser usada para separar correctamente classes que não são linearmente separáveis. No entanto, isto pode levar a sobreajustamento (*overfitting*). Explique isto focando os seguintes aspectos por ordem decrescente de prioridade:

- Porque é que a Regressão Logística é uma regressão e qual o valor que tenta prever.
 - Como se pode usar essa previsão para usar a Regressão Logística como um classificador.
 - O que se tem de fazer para poder separar classes, que não são linearmente separáveis, com este classificador.
- O que é o sobreajustamento (overfitting) e como se pode tentar mitigar isto na Regressão Logística assumindo que o modelo já está determinado (não se pode escolher outro modelo).*

English:

Attention: your answer should be concise and well structured. If you feel you do not have time to give a well structured answer covering all items below focus on the most important ones.

Logistic Regression is linear classifier but it can be used to separate classes that are not linearly separable. However, this can lead to overfitting. Explain this focusing on the following aspects, in decreasing order of importance:

- Why Logistic Regression is a regression model and what value does it try to predict.
- How can we use this prediction to use Logistic Regression as a classifier.
- What do you need to do to separate classes, that are not linearly separable, with this classifier.
- What is overfitting and how can you mitigate it with Logistic Regression assuming that the model has already been determined (you cannot choose another model).

Pergunta 3

Português

Atenção: a sua resposta deve ser concisa e bem estruturada. Se vir que não tem tempo para dar uma resposta estruturada que cubra tudo o que é pedido, foque apenas os itens mais importantes.

Explique o que há de semelhante e diferente entre o Classificador de Bayes e o Classificador *Naïve* Bayes focando os seguintes aspectos por ordem decrescente de prioridade:

- Em que se baseiam para atribuir cada exemplo a uma classe.
- O que é que o Classificador *Naïve* Bayes assume para simplificar este cálculo.

- Porque é que há vantagem em fazer a simplificação que o Classificador *Naïve Bayes* faz.
- Explique como podemos criar um Classificador *Naïve Bayes* paramétrico e um Classificador *Naïve Bayes* não-paramétrico.

English:

Attention: your answer should be concise and well structured. If you feel you do not have time to give a well structured answer covering all items below focus on the most important ones.

Explain the similarities and differences between the Bayes Classifier and the Naïve Bayes Classifier focusing on the following aspects, in decreasing order of importance:

- What do these classifiers use to assign each example to a class.
- What does the Naïve Bayes Classifier assume to simplify this computation.
- What is the advantage of doing the simplification that the Naïve Bayes Classifier does.
- Explain how we can create a parametric Naïve Bayes Classifier and a non-parametric Naïve Bayes Classifier.

Pergunta 4

Português

Atenção: a sua resposta deve ser concisa e bem estruturada. Se vir que não tem tempo para dar uma resposta estruturada que cubra tudo o que é pedido, foque apenas os itens mais importantes.

A Máquina de Vectores de Suporte (SVM) é um classificador versátil e eficiente. Explique o que é e quais as suas vantagens focando os seguintes aspectos por ordem decrescente de prioridade:

- O que são vectores de suporte e onde é que a SVM tenta colocar o discriminante.
- O que é e o faz o parâmetro C.
- O que é o *kernel trick* e como permite classificar correctamente exemplos quando as classes não são linearmente separáveis.
- Explique porque é que o uso do *kernel trick* pode agravar o sobreajustamento (*overfitting*).

English:

Attention: your answer should be concise and well structured. If you feel you do not have time to give a well structured answer covering all items below focus on the most important ones.

The Support Vector Machine (SVM) is a versatile and efficient classifier. Explain what it is and what are its advantages focusing on the following aspects, in decreasing order of importance:

- What are support vectors and where does the SVM try to place the discriminant.
- What is the parameter C and what it does.
- What is the kernel trick and how it allows the SVM to correctly classify examples when the classes are not linearly separable.

- Explain how using the kernel trick can lead to more overfitting.

Pergunta 5

Português

Atenção: a sua resposta deve ser concisa e bem estruturada. Se vir que não tem tempo para dar uma resposta estruturada que cubra tudo o que é pedido, foque apenas os itens mais importantes.

Bagging e *AdaBoost* são dois métodos usados para agregar hipóteses de forma a mitigar alguns problemas na previsão. Explique estes métodos, as suas vantagens e em que situações devem ser usados focando os seguintes aspectos por ordem decrescente de prioridade:

- Como é que se obtém as diferentes hipóteses em *Bagging* e em *AdaBoost*.
- Qual dos dois é mais adequado para usar com modelos instáveis, com muita variância, e porquê.
- Porque é que o outro é adequado para modelos com um *bias* elevado.

English:

Attention: your answer should be concise and well structured. If you feel you do not have time to give a well structured answer covering all items below focus on the most important ones.

Bagging and *AdaBoost* are two methods used to aggregate hypotheses in order to mitigate some problems with prediction. Explain these methods, their advantages and in which situations they should be used focusing on the following aspects, in decreasing order of importance:

- How do we obtain different hypothesis in *Bagging* and in *AdaBoost*.
- Which of the two methods is more adequate for using with unstable methods, with high variance, and why.
- Why is the other method good for using with models with a high bias.