

Computer Vision Exam**Name:** Yasuo Ignacio Maidana Pérez**Student ID:** A01328427**All questions are worth 5 points, except for 18 which is worth 10 points.****1. Briefly describe the 7 steps of machine learning.** ✓

1. Recollection of data. For any machine learning application, it is mandatory first have data, because machine learning uses statistics to generate a statistic model to predict the system's behavior
2. Data processing. Although it is necessary a good amount of data, it will be useless we ensure we have not only valid data (not crashed or incomplete) but also representative. For this reason, we need to manipulate and refine data.
3. Model selection. We need to define, the type of model, if it will be categorical or regressive, supervised, unsupervised, semi supervised... etc.
4. Model training. Once we have decided the model and what data to use, in this step the model is generated by a heuristic method that optimizes some values based on some assumption
5. Model evaluation. The model is evaluated in order to determine if it is appropriate or not. In this part we must be very cautious avoiding overfitting or underfitting
6. Model tuning. Sometimes some values must be manipulated by the engineer or scientist in order to get a better performance
7. Prediction. Finally, we can apply our model into the environment

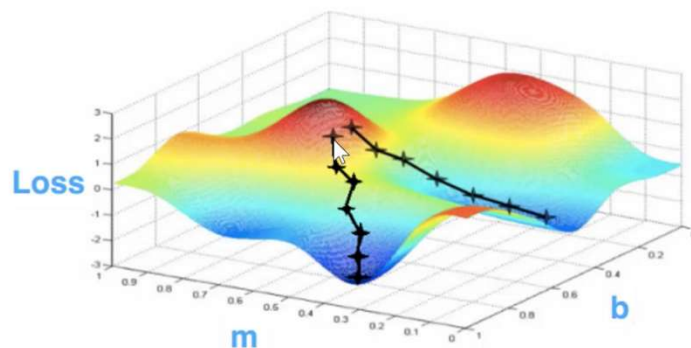
2. Explain the concept of gradient descent algorithm, use a diagram to support your answer and explain the use of all the terms and constants that you introduce. ✓

The gradient descent algorithm is based on the idea of partial derivatives. If we remember, the through the gradient we can find the path that give us to the lowest point in a function. Here, through iteration, the computer finds the minimum values that makes our equation (model) behave the most optimal way. For a linear regression, the model finds two parameters, in class we defined them as h_0 and h_1 (we refer them as parameters), there exist two other parameters that are defined by ourselves (the other ones are calculated by computing), those values are the

number of iterations, the “size” of the steps and the bias (these are the hyperparameters)

Gradient Descent

$$f(x) = \text{nonlinear function of } x$$

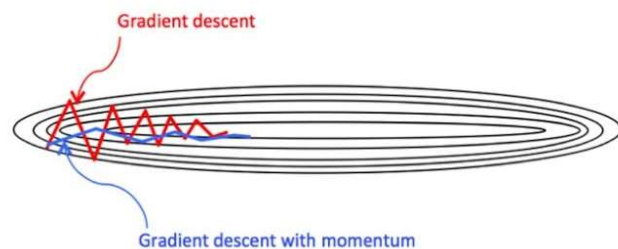
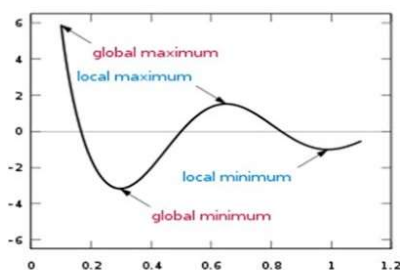


3. Briefly describe the stochastic gradient descent method. What are its advantages and disadvantages compared to gradient descent? ✓

The stochastic gradient descent is a method very similar to GD, however instead of using all the data available, we only use a point (or a set of data called batches). The advance of this method is that we reduce for much the computing power needed to do the task. However, since it doesn't use all the data, this strategy tends to be very sensitive to noise, and its design stability relies a lot on data's distribution

4. Explain the use of momentum in gradient descent, accompany your explanation with a diagram. Explain the use of all the terms and constants you introduce in your answer. ✓

The momentum gradient is a computer optimization method that through iterations allows us to find the point where a function reach its lowest point. It uses it derivative as GD, however the difference relies on that this method “remembers” how the past values affected the perform of the values founded. It “punishes” based on previous values through its partial derivative, this value is calculated using the hyperparameter alfa (size of step), and the hyperparameter up to square



5. What is overfitting and underfitting? ✓

Both are error of our models, however underfitting refers to a model that is over simplified, so it can't predict well futures values. On the other hand, we reach overfitting when our model almost (or) memorize the data, so when we use it with previous data, it seems that works very well, however it has a lot of problem predicting new values

6. Explain the bias-variance tradeoff. ✓

It is a kind of disjunction between how precise or accurate we want our model. Because on most of the time, we must choose one of them or we couldn't reach any model

7. Explain how to identify overfitting in a model. ✓

Comparing the training error variance to validation error variance, if the difference is high, there must exist an overfitting. Identifying it is a big challenge now of design a new model

8. Suppose your model is overfitting. Which of the following is NOT a valid way to try and reduce the overfitting?

1. **Increase the amount of training.**
2. **Improve the optimization algorithm being used for error minimization.**
3. **Decrease the model complexity.**
4. **Reduce the noise in the training data.**

9. A student created a model that predicts house pricing based on the house's color and area. To handle the categorical data (house's color), he assigned the following numbers to each color:

White = 1, Blue = 2, Red = 3.

What is wrong with this approach and how would you fix it? Show how would you handle the categorical data.

The error relies that the student is giving them an ordinal value, so the computer could “think” that white+blue=red, when it doesn’t have sense. An approach to fix it could be through matrix representation, making the colors a kind of flags that “turn on or off” the parameters as the situation requires

10. How would you screen for outliers and what should you do if you find one? ✓

We can find them using a boxplot data representation, they are located far away from the “average/normal” distribution. If I find one, before removing it, I will find its causes and see if there exist an important relationship with the environment studied

11. You are responsible for generating an image classifier. Your model should be able to classify different types of animals. Please explain two approaches you would use to collect large volumes of labeled images. ✓

One of them could be crowdsourcing, this way relies on worldwide (or local, it will depend on the tool) contribution made by other people.

Web-scraping, here we will use an algorithm to download the

12. What is the output of the following python code? ✓

```
1. foobar = { 'title' : ['Hitchhiker', 'Guide', 'Galaxy'],
2.            'author' : 'Douglas Adams',
3.            'numbers' : [[32,33,33,30,31,32,35],
4.                          [28,29,29,27,29,26,28],
5.                          [41,40,44,44,41,42,43],
6.                          [37,38,39,33,31,33,37]]}
7.
```

```
8. print(foobar['numbers'][2][-2])
```

Imprime el valor localizado en la fila 2 y antepenultima columna de la matriz 'numbers'

42

13. What is the output of the following python code? ✓

```
1. foobar = {'s1' : 'Illustrate your point with an example',
2.          's2' : 'Minatitlan is 5 miles away',
3.          's3' : 'TI? It stands for Texas Instruments'}
4.
5. for foo in foobar:
6.     print(foobar[foo][:4])
```

Imprime los primeros 4 caracteres de las oraciones guardadas en el diccionario

Illu

Mina

TI?

14. How does the bias-variance decomposition of a ridge regression estimator compare with that of ordinary least square regression? (Select one)

- A) Ridge has larger bias, larger variance.
- B) Ridge has smaller bias, larger variance.
- C) Ridge has larger bias, smaller variance.
- D) Ridge has smaller bias, smaller variance.



15. Compare and contrast K-fold cross-validation and hold out approach. What are the advantages and disadvantages of each method? ✓

The advantage of k-fold over hold out relies on the “efficiency” on manipulating data, since it uses all of them. However, since it ‘cut’ the data in segments, it needs to compute for every batch, so using k-fold method implies more time

16. Which of the following options is/are true for K-fold cross-validation? ✓

- 1. Increase in K will result in higher time required to cross validate the result.
- 2. Increase in K will result in lower time required to cross validate the result.
- 3. K can be any real number.
- 4. If K=N, then it is called Leave one out cross validation, where N is the number of observations

17. Assume you are the chief editor of the *International Journal on Artificial Intelligence* and decided not to publish a submission with the following claim:

“My method achieves a training error lower than all previous methods!”

How would you justify your decision of discarding this submission?

There could exist several reasons

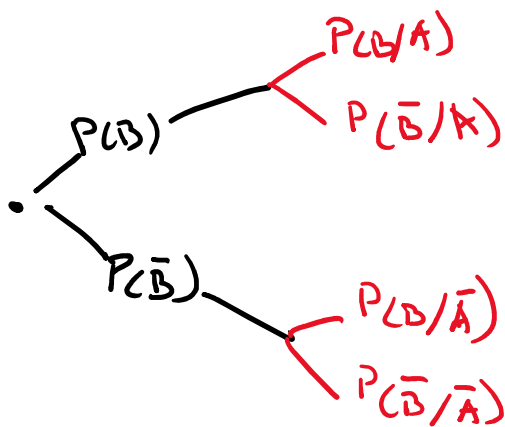
1. The model proposed in fact has a lower error, but for specified scenarios
2. The model wasn't compared
3. The number of tests is not reliable
4. The number of iterations is not reliable



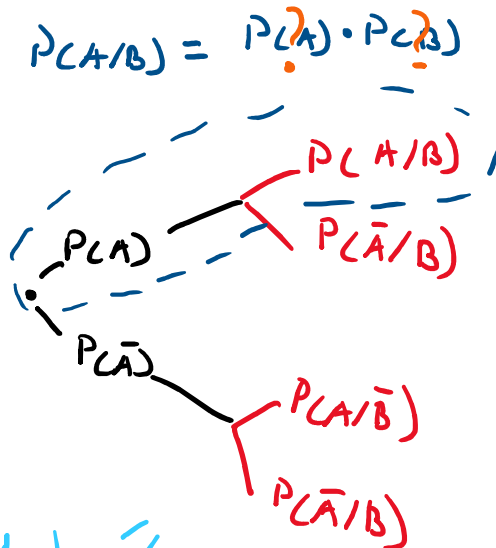
18. Based on the following fact:

$$P(A|B) = 2/3$$

a) Do you have enough information to compute $P(B|A)$? If not, write "not enough data". If so, compute the value of $P(B|A)$. ✓



"Not enough data"

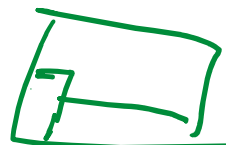
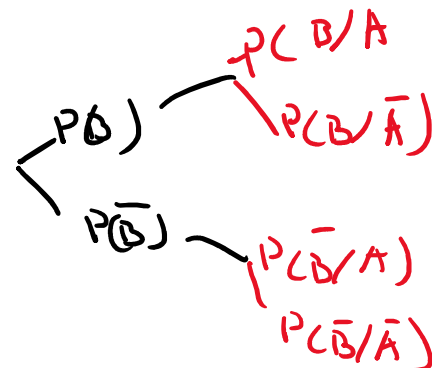
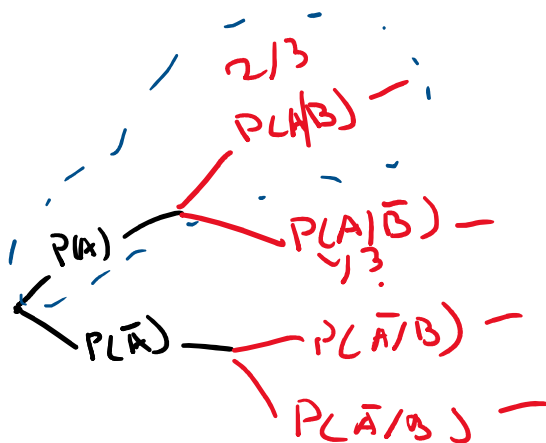


b) Given this additional data:

$$P(A|B) = 2/3$$

$$P(A|\sim B) = 1/3$$

$$P(A) \cdot \frac{2}{3} = P(B) \cdot P(B|A)$$



No sabemos si son complementarias

Do you have enough information to compute $P(B|A)$? If, not write "not enough data". If so, compute the value of $P(B|A)$.

Not enough ✓

c) Given this additional data:

$$P(A|B) = 2/3$$

$$P(A|\sim B) = 1/3$$

$$P(B) = 1/3$$

Do you have enough information to compute $P(B|A)$? If, not write "not enough data". If so, compute the value of $P(B|A)$.

$$P(\bar{B}) = 2/3$$

$$P(B|A) = \frac{P(B) P(A|B)}{P(B) P(A|B) + P(\bar{B}) P(A|\bar{B})}$$

$$P(B|A) = \frac{1}{2}$$

d) Given this additional data:

$$P(A|B) = 2/3$$

$$P(A|\sim B) = 1/3$$

$$P(B) = 1/3$$

$$P(A) = 4/9$$

Do you have enough information to compute $P(B|A)$? If, not write "not enough data". If so, compute the value of $P(B|A)$.

$$P(B|A) = \frac{P(B) P(A|B)}{P(B) P(A|B) + P(\bar{B}) P(A|\bar{B})}$$

same as above

$$P(B|A) = \frac{1}{2}$$

19. Explain a concept of ML that was not included in this exam that you learned in this first week. Show me what you have learned!

Logistic regression. In this technique we apply regression methods to a probability function in order to find relationships between some elements and a characteristic.

Bonus

You can each earn some extra credit on your exam. You get to choose whether you want 5 points added to your grade, or 15 points. But there's a catch: if more than 10% of the class selects 15 points, then no one gets any points. All selections are anonymous.

%15

Buen examen Yasuo! Me gustó leer tus respuestas, principalmente las relacionadas a GD y sus variantes, bastante bien explicado cada uno de los métodos. Otro aspecto que resalto es que te preparaste bastante bien para el problema de Bayes! Tienes muy buenas bases de programación, matemáticas y ML, sigue con este buen trabajo.