**COSC 6335**

**Group H Homework Group Credit**

**To be presented October 22**

Compute the GINI-gain[1] and information gain[2] for the following decision tree split:

Node N (5,5,5,5) “the node contains 5 examples of for classes C1, C2, C3 and C4 each”

and is splitt using a 3-way split into node N1, N2, N3 with the following class proportions:

N1=(5,0,0,0)

N=(5,5,5,5) N2=(0,0,5,4)

N3=(0,5,0,1)

**Homework Group Credit Task COSC 6335**

**Prepare a Tensorflow Playground Demo**

**Group I**

**to be presented on Tuesday, October 29**

Each of you initially plays around with the tool <http://playground.tensorflow.org/> centering on classification sets in the tool; however,

* Explore the impact the learning rate has on the learning process in general and the final result
* Expose the datasets to noise up to 40% and analyze what the impact the noise has on obtained training and test loss and observe if overfitting occurs.
* Explore the impact of using L1 regularization for the original datasets and on datasets with 10% and 30% noise
* If you can, show some convincing examples of overfitting

Next prepare a 8-12 minute Tool demo of the most interesting runs you observed. Finally, prepare a summary slide (or 2, if you need more space) of what you observed when explored the tool and discuss the slide in 2-4 minutes after your demo.

Group Homework Credit

Group J Task

Neural Networks

*To be presented during the online class on Nov. 1*

a) Assume the following subset of a neural network is given:

wA,B=0.2

ΔB=0.5

A B

Assume that B is an intermediate node of a neural network, the forward propagation activation values of nodes aA and aB are 0.5 and 0.8 and the current value of WA,B is 0.2; the associated error ΔB of node B that was computed by the back propagation algorithm is 0.4, the learning rate γ is assumed to be 0.5. First give the general weight update formula and then compute the new value of weight wA,B!

b) The step size in neural network learning depends—among other factors—on the gradient of the error function[[1]](#footnote-1); explain why this is important!

c) Take a look at the sub neural network consisting of nodes A, B, C, and D in the figure below; give a formula that computes the associated error ΔA for a node A. Assume the used activation function is g and its derivative is denoted by g’, and the activation of a node X is denoted by aX and the linear input of a node X is denoted by zX. First provide a general formula; then, replace general variables in the formula by their actual known values.

wA,B=0.2

ΔB=0.4

A B

wC**,A**=1 wD,A=0.5

C D

Group Homework Credit

Group K Task

Ethics of Data Science Discussion

to be take place during the class on Thursday, November 7

[**Ethical aspects for data science include1**](https://www.bing.com/ck/a?!&&p=aef8b00c9face557ad12b2615289ebb8c2425a05122078ee2c8bb21452758dbdJmltdHM9MTcyOTU1NTIwMA&ptn=3&ver=2&hsh=4&fclid=113a146f-7b4b-6d4f-1e6d-046a7aad6cfc&psq=Ethical+Aspects+of+Data+Science&u=a1aHR0cHM6Ly93d3cuYW5hbHl0aWNzdmlkaHlhLmNvbS9ibG9nLzIwMjIvMDIvZXRoaWNzLWluLWRhdGEtc2NpZW5jZS1hbmQtcHJvcGVyLXByaXZhY3ktYW5kLXVzYWdlLW9mLWRhdGEv&ntb=1)[**2**](https://www.bing.com/ck/a?!&&p=22e5ccc34412cc5211a1974b101af374b7f0b65a7deff527e8f1b95105e6b144JmltdHM9MTcyOTU1NTIwMA&ptn=3&ver=2&hsh=4&fclid=113a146f-7b4b-6d4f-1e6d-046a7aad6cfc&psq=Ethical+Aspects+of+Data+Science&u=a1aHR0cHM6Ly9lbWVyaXR1cy5vcmcvYmxvZy9kYXRhLXNjaWVuY2UtYW5kLWFuYWx5dGljcy1kYXRhLXNjaWVuY2UtY291cnNlLWN1cnJpY3VsdW0v&ntb=1):

* Transparency
* Accountability
* Fairness
* Privacy
* Security
* Consent
* Integrity

[Data scientists should be aware of potential biases in their analyses and ensure that individuals from various demographic groups are treated fairly and equally**2**](https://www.bing.com/ck/a?!&&p=22e5ccc34412cc5211a1974b101af374b7f0b65a7deff527e8f1b95105e6b144JmltdHM9MTcyOTU1NTIwMA&ptn=3&ver=2&hsh=4&fclid=113a146f-7b4b-6d4f-1e6d-046a7aad6cfc&psq=Ethical+Aspects+of+Data+Science&u=a1aHR0cHM6Ly9lbWVyaXR1cy5vcmcvYmxvZy9kYXRhLXNjaWVuY2UtYW5kLWFuYWx5dGljcy1kYXRhLXNjaWVuY2UtY291cnNlLWN1cnJpY3VsdW0v&ntb=1). [They should also protect privacy and confidentiality, assess societal impact, and respect intellectual property**2**](https://www.bing.com/ck/a?!&&p=22e5ccc34412cc5211a1974b101af374b7f0b65a7deff527e8f1b95105e6b144JmltdHM9MTcyOTU1NTIwMA&ptn=3&ver=2&hsh=4&fclid=113a146f-7b4b-6d4f-1e6d-046a7aad6cfc&psq=Ethical+Aspects+of+Data+Science&u=a1aHR0cHM6Ly9lbWVyaXR1cy5vcmcvYmxvZy9kYXRhLXNjaWVuY2UtYW5kLWFuYWx5dGljcy1kYXRhLXNjaWVuY2UtY291cnNlLWN1cnJpY3VsdW0v&ntb=1).

Pick a Data Science Ethics Discussion Topic. Introduce the topic by presenting 3-4 slides. Then lead the in-class discussion!

1. Also called loss or cost function in some textbooks. [↑](#footnote-ref-1)