Data Science Quiz

Question 1

Which of the following methods are supervised / unsupervised? *(10 points)*

method	supervised	unsupervised
Linear Regression	✓	
Support Vector Machines	✓	
Variational Autoencoder		✓
Multi-Layer Perceptron	✓	
Principal Component Analysis		✓
Gaussian Mixture Models		✓
Gradient Boosting Trees	✓	
Naive Bayes	✓	
NMF		✓
k-Means clustering		✓

Question 2

Sort the following Machine Learning methods by age: *(5 points)*

Logistic Regression, Neural Networks, Random Forest, Linear Regression, SVM

1.	Linear Regression: Beginning of 19th century
2.	Logistic Regression: 1933; Different sources give different dates
3.	Neural Networks: 1943
4.	SVM: 1963
5.	Random Forest: End of 1990's beginning of 2000's

Question 3

Who are the persons in the pictures? *(8 points)*

a)

b)

c)

d)

Question 4

Find 10 bugs: *(10 points)*

```
for year in range(1890, 2015, 1):
    total = 0
    filename = 'names\yob{year}.txt'
    for line in open(filename, 'w'):
        columns = line.split(',').strip()
    total += float(columns[2])

print("Result: {} births total".format(year))
```

total

Question 5

What do the following bash commands do? *(6 points)*

`ls -a`	It lists all files including hidden ones
`sudo rm -rf /`	Delete everything from your root directory
`chmod 700 *`	rwx permission on all files of the directory for the owner and no permissions for the group and others
`grep print *.py | wc -l`	Counts the number of lines containing the word „print“ in all python files of the current directory

Question 6

Name three hyperparameters of a Random Forest. *(3 points)*

- max depth
- max features
- n_estimators

Question 7

Write an SQL query that extracts the 10 most frequently occurring items in the *‘subject’* column from the table *‘data_scientists’*, but only consider students with the column *’python’* being 1 or higher. Output results in descending order. *(10 points)*

```
SELECT subject, count(subject) as value
FROM data_scientists
WHERE python => 1
GROUP BY subject
ORDER BY value DESC
LIMIT 10;
```

Question 8

Name the functions. *(12 points)*

$P(A\|B)=\frac{P(B\|A)P(A)}{P(B)}$	Bayes Theorem
$\frac{1}{N}\sum_i (y_i - y_i^{true})^2 + \lambda \sum_j b_j^2$	MSE + L2 regularization
$\frac{1}{1+e^{-x}}$	Sigmoid Function
$\frac{e^{z_i}}{\sum_j e^{z_j}}$	Softmax

Question 9

Write one item you could import from each Python module. *(4 points)*

`pandas`	`DataFrame`
`random`	rand
`numpy`	array(…), arange
`seaborn`	lineplot
`os`	listdir

Question 10

Which strings does the Regular Expression '`R[oau]\w+e`' match? *(8 points)*

Rome	~~rose~~	~~Rue~~	~~Dome~~
~~Rhizome~~	~~Rhizome~~	~~Ru\w+e~~	Raave

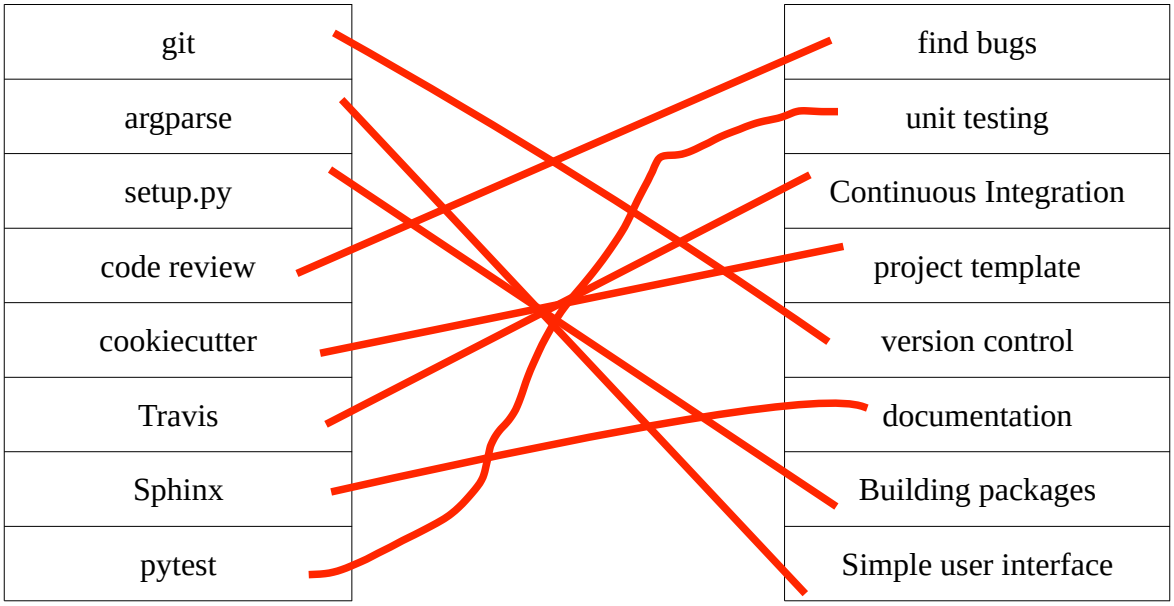
Question 11

Name five Python Exceptions (e.g. `SyntaxError`). *(5 points)*

- ValueError
- KeyError
- EOF
- ZeroDivisionError
- KeyboardInterruptError

Question 12

Match pairs. *(8 points)*



© 2018 SPICED Academy.
May be re-used for non-commercial use only.
www.spiced-academy.com