Python: Annotation

Blank notebook to be used for class exercises.

Exercise 1

Complete the Sentiment Annotation Survey on Blackboard.

Exercise 2

Judge the annotations in following Data:

sentiment.txt

You are the judge. Look over the annotations and come up with the following:

- What is your judgment for the correct entity + sentiment annotation?
- How would you amend the annotation guidelines to solicit more consistent annotations?

While judging the annotations, keep track of them below:

1. **Target:** ISIS, **Sentiment:** Negative
2. **Target:** US Senate , **Sentiment:** Mixed
3. **Target:** Major Meeting, **Sentiment:** Unknown
4. **Target:** Sean Spicer, **Sentiment:** Mixed
5. **Target:** The West, **Sentiment:** Positive
6. **Target:** Fake News Media, **Sentiment:** Negative
7. **Target:** Obamacare Victims, **Sentiment:** Negative
8. **Target:** Ivanks, **Sentiment:** Positive
9. **Target:** ISIS, **Sentiment:** Positive
10. **Target:** Emmanuel Macron, **Sentiment:** Positive
11. **Target:** VA State Police, **Sentiment:** Positive
12. **Target:** Sailors, **Sentiment:** Positive
13. **Target:** Fake News Story, **Sentiment:** Negative

14. **Target:** President Trump, **Sentiment:** Unknown
15. **Target:** Democrat Jon Ossof, **Sentiment:** Negative
16. **Target:** Charlottesville, **Sentiment:** Negative
17. **Target:** Joint Base Andrews , **Sentiment:** Unknown
18. **Target:** Brave Men and Women , **Sentiment:** Positive
19. **Target:** America , **Sentiment:** Positive
20. **Target:** #NABTU2017, **Sentiment:** Positive

How many of your judgements match the majority view?

Exercise 3

Manually calculate Cohen's kappa for the tables in the slides and put your answers below:

Answers here

In [8]:
```
#kappa question 1
po_1 = (25 + 25)/100
pe_1 = (50/100) * (50/100) + (50/100) * (50/100)
numerator_1 = po_1 - pe_1
denominator_1 = 1 - pe_1
kappa_1 = numerator_1 / denominator_1

#kappa question 2
po_2 = (50 + 50)/100
pe_2 = (50/100) * (50/100) + (50/100) * (50/100)
numerator_2 = po_2 - pe_2
denominator_2 = 1 - pe_2
kappa_2 = numerator_2 / denominator_2

#kappa question 3
po_3 = (0 + 0)/100
pe_3 = (50/100) * (50/100) + (50/100) * (50/100)
numerator_3 = po_3 - pe_3
denominator_3 = 1 - pe_3
kappa_3 = numerator_3 / denominator_3
```

In [9]:
```
print(kappa_1)
print(kappa_2)
print(kappa_3)
```

```
0.0
1.0
-1.0
```

Exercise 4

Write code to calculate and print the cohen's kappa between rater1 and rater2.

Hint: You will need to create the confusion matrix (matrix of how many times each rater agrees for each item)

In [12]:
```
rater1 = ['yes','no','yes','yes','yes','yes','no','yes','yes']
rater2 = ['yes','no','no','yes','yes','yes','yes','yes','yes']
```

In [13]:
```
r1_yes_r2_yes = 0
r1_no_r2_no = 0
r1_no_r2_yes = 0
r1_yes_r2_no = 0

for r1, r2 in zip(rater1, rater2):
    if r1 == 'yes' and r2 == 'yes':
        r1_yes_r2_yes += 1
    elif r1 == 'no' and r2 == 'no':
        r1_no_r2_no += 1
    elif r1 == 'no' and r2 == 'yes':
        r1_no_r2_yes += 1
    else:
        r1_yes_r2_no += 1
```

In [15]:
```
print(r1_yes_r2_yes, r1_no_r2_no, r1_no_r2_yes, r1_yes_r2_no)
```

```
6 1 1 1
```

In [16]:
```
po = (r1_yes_r2_yes + r1_no_r2_no) / (r1_yes_r2_yes + r1_no_r2_no + r1_no_r2_yes + r1_yes_r2_no)
p_r1_yes = (r1_yes_r2_yes + r1_yes_r2_no) / (r1_yes_r2_yes + r1_no_r2_no + r1_no_r2_yes + r1_yes_r2_no)
p_r1_no = (r1_no_r2_no + r1_no_r2_yes) / (r1_yes_r2_yes + r1_no_r2_no + r1_no_r2_yes + r1_yes_r2_no)

p_r2_yes = (r1_yes_r2_yes + r1_no_r2_yes) / (r1_yes_r2_yes + r1_no_r2_no + r1_no_r2_yes + r1_yes_r2_no)
p_r2_no = (r1_no_r2_no + r1_yes_r2_no) / (r1_yes_r2_yes + r1_no_r2_no + r1_no_r2_yes + r1_yes_r2_no)

pe = (p_r1_yes * p_r2_yes) + (p_r1_no * p_r2_no)

kappa = (po - pe) / (1 - pe)
```

In [17]:
```
print(kappa)
```

```
0.35714285714285715
```

In [ ]:
```

```