Problem Set 05

1. Winner (4p)

After 931.554 seconds, Han won a 10 round race against Luke (who finished 2.802 seconds later) and Nien (who finished with an additional 0.928 seconds). The two losers claim that there is no statistical significant difference between Han and them at the 5% level and demand a rematch. Are they correct and should Han race once more against both of them? In the following table you find the individual lap times in seconds. Use the Student's t-test to check the difference between Han-Luke, Han-Nien, and Luke-Nien. Note your observations. You can use a spreadsheet program but without the built-in t-test formula.

Lap	Han	Luke	Nien
01	92.704	93.440	93.866
02	92.800	93.266	92.700
03	92.757	93.231	92.987
04	93.863	93.334	93.152
05	93.001	93.449	93.307
06	93.026	93.468	94.237
07	93.052	93.425	93.771
08	93.185	93.367	93.441
09	93.309	93.475	93.125
10	93.857	93.900	94.698

2. Internet (6p+2p)

The main goal of a given system is to retrieve only the very cute photos of cats out of the total 54 cat pictures. We know that there are 10 images in the collection that are very cute and the rest is just cute. The system manages to return 8 very cute cat pictures but also retrieves 10 just cute images. Create the confusion matrix and calculate the overall precision, recall, and F1 of this system.

The 18 retrieved pictures are returned in the following order (the leftmost image is at the first rank; VC = very cute; JC = just cute):
VC, VC, JC, VC, JC, VC, VC, VC, VC, JC, JC, JC, JC, JC, JC, VC, JC, JC

Calculate precision and recall at each rank (so, first with "VC", then "VC, VC", then "VC, VC, JC", and so on). You don't need to create the confusion matrix each time.

Bonus: There is a balance point when precision and recall are the same. Must there always be such a point between precision and recall? What is the relationship between the F1 value and this balance-point? (+2p)

Deadline:

November 7, 2016 at 8:00 AM