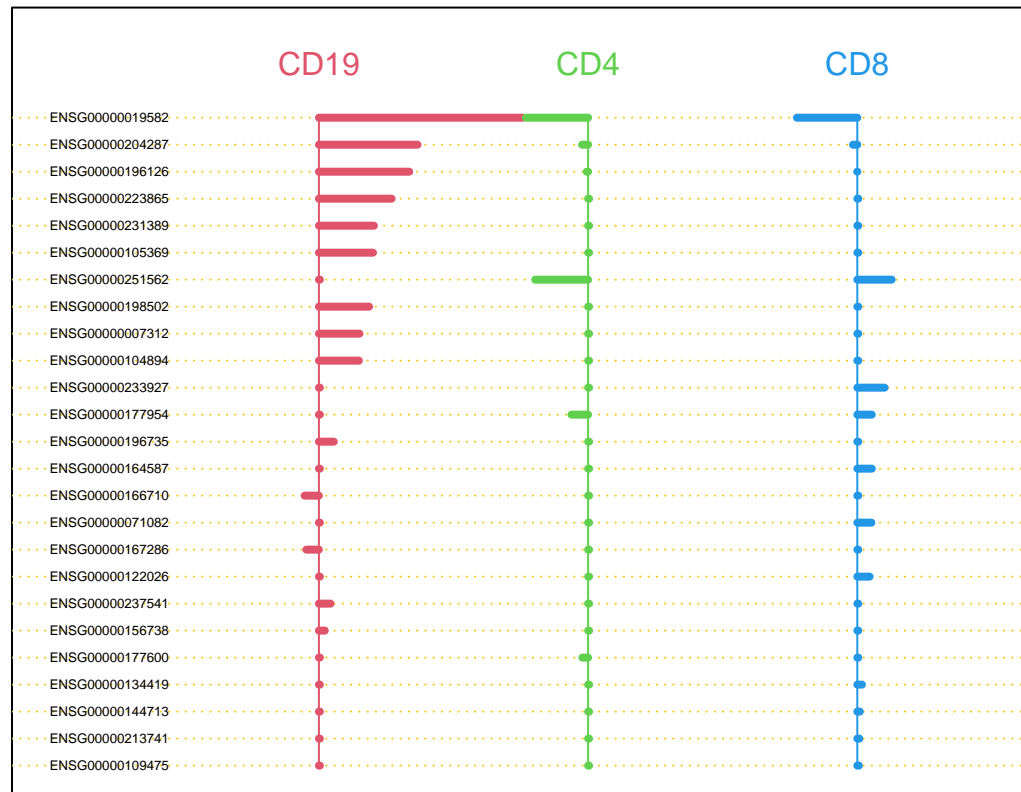


Assignment 3

Martynas Lukosevicius

15/11/2020

1.



Centroid plot shows distance from gene mean value to global mean value. Positive or negative shows if mean is greater or smaller than the global mean. No, it can't happen that all values for some gene in centroid plot are positive. Number of genes selected when threshold = 7.212944 : 25.

2.

Most contributed genes: ENSG00000019582, ENSG00000204287. alternative names: CD74, HLA-DRA. These genes are marker genes for B cells

Test error rate: 0.1111111

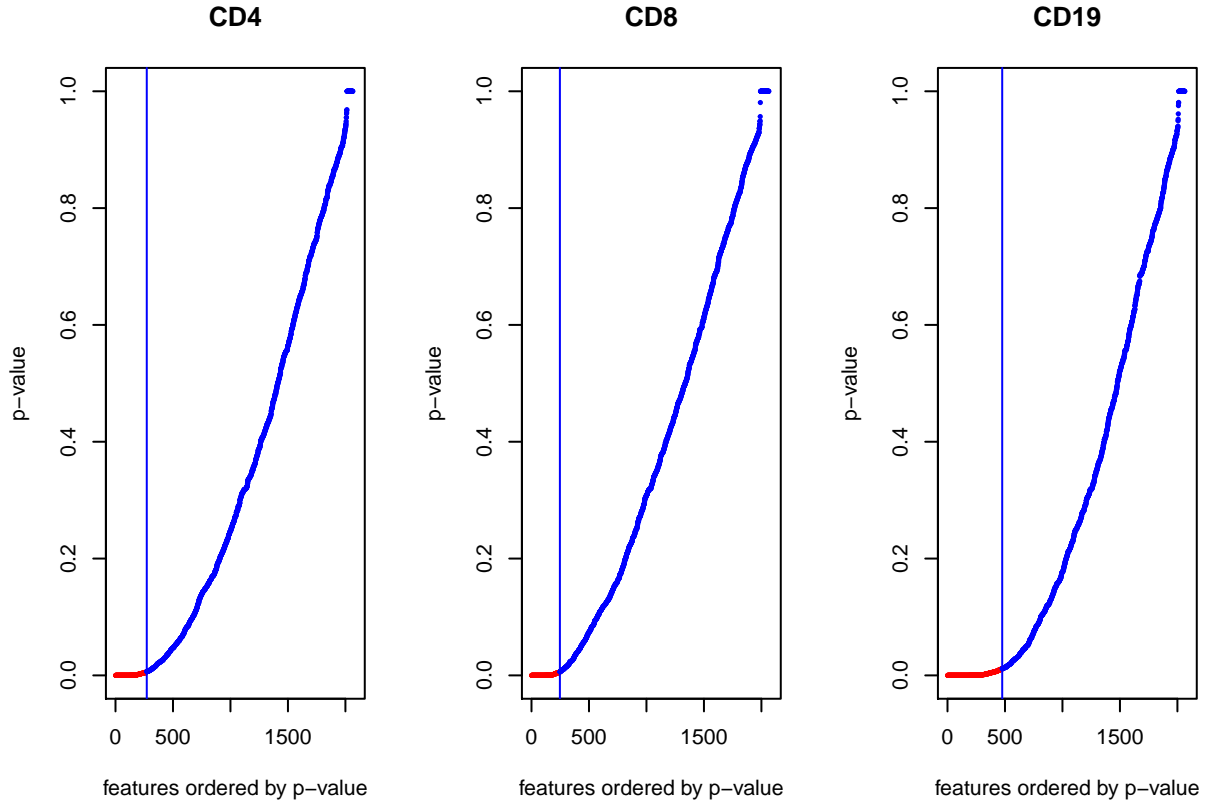
3.

Table below shows test error rate and number of features used.

	Test error rate	N. features
NSC	0.110	25
EN	0.056	143
SVM	0.022	2085

I prefer Elastic net, because SVM might overfit and use a lot of parameters, NSC might have underfitted.

4.



points symbolize genes, red genes are rejected meaning that they have an effect on corresponding cell.

Table below shows the amount of rejected genes for each cell.

cell	rejected
CD4	272
CD8	247
CD19	476