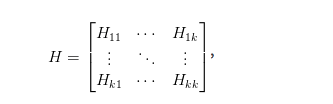
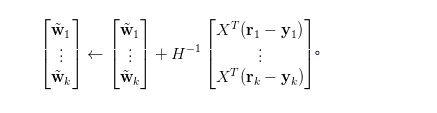
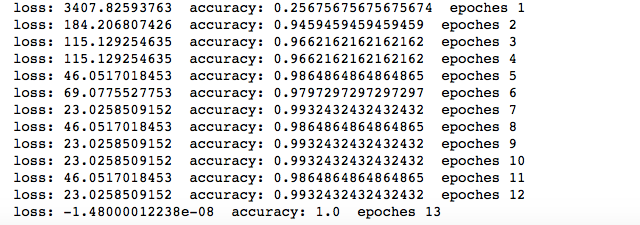
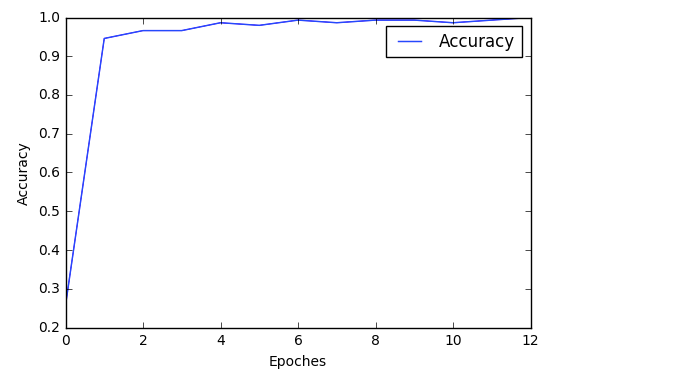
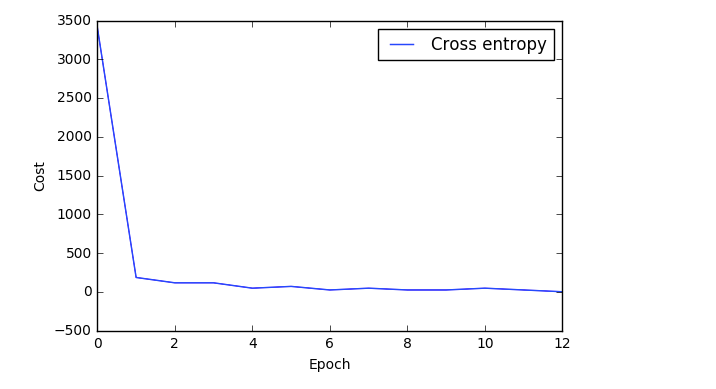
3.1 :

For Hessians , it is diagonal :

For update W, it needs to update Hessian inverse dot X.T(truth class - prediction class):

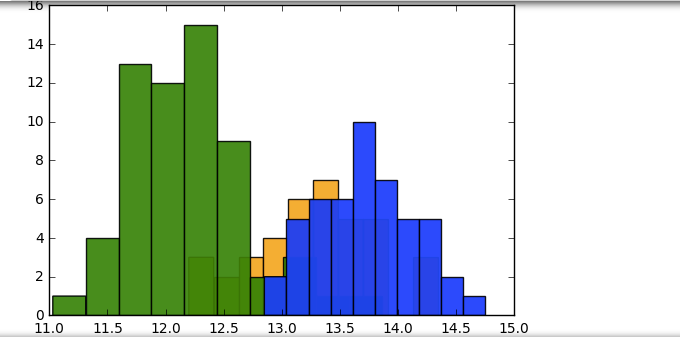


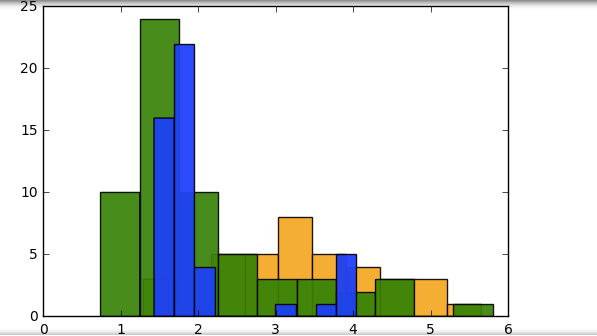
As we can see for hessians it can converges quickly in less epochs.

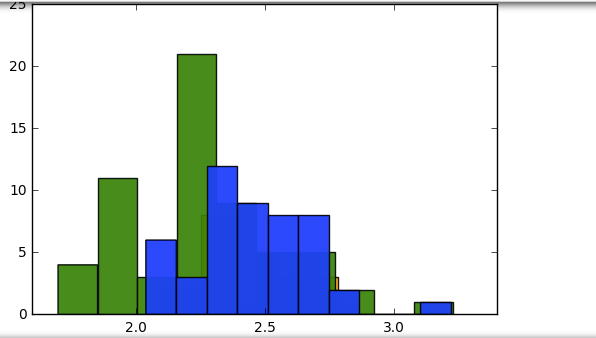


3.2

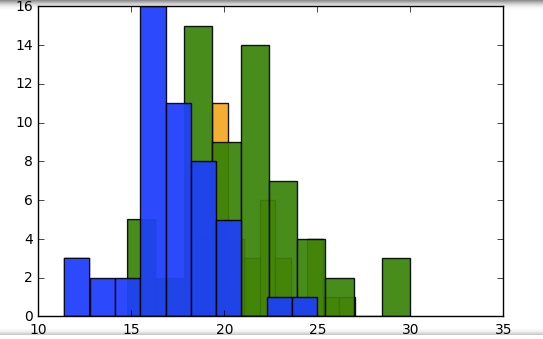
3.3

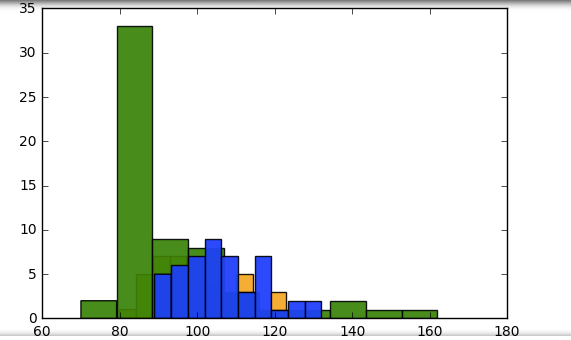
Variable 1:

Variable 2:

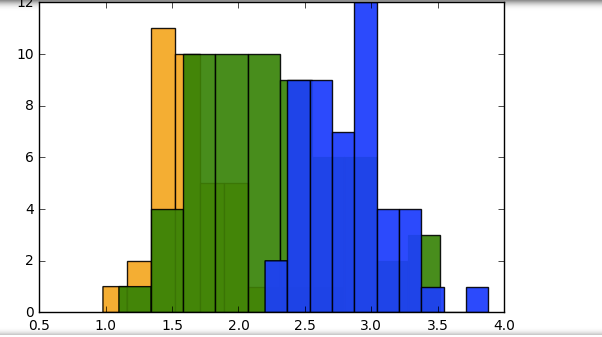
Variable 3:

Variable 4:

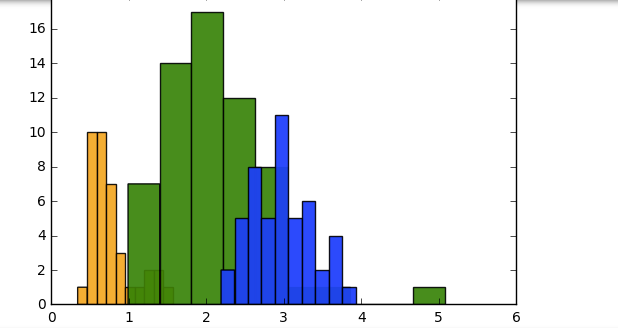


Variable 5:

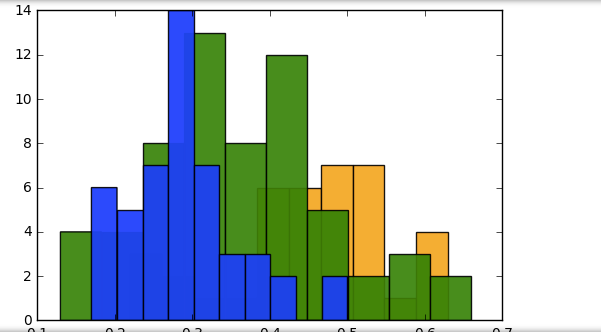
Variable 6:



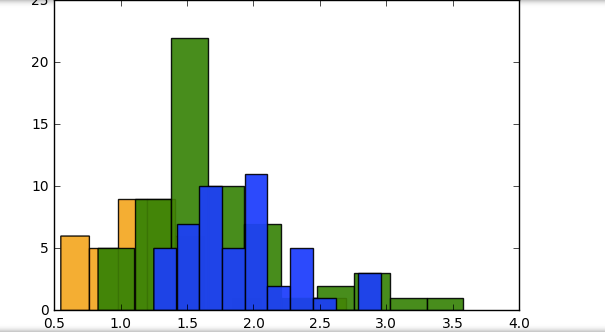
Variable 7:



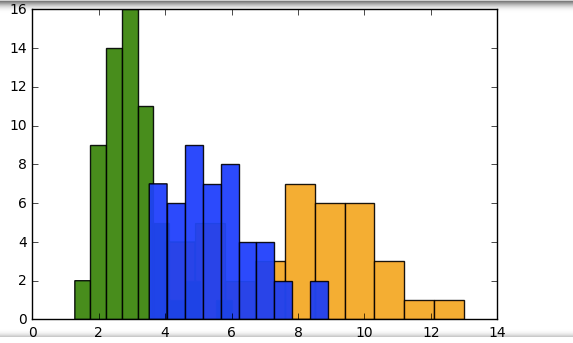
Variable 8:

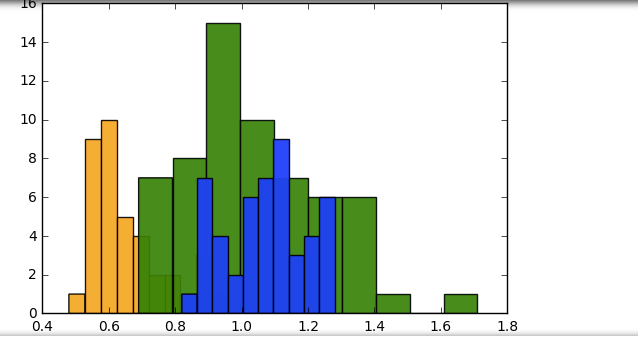


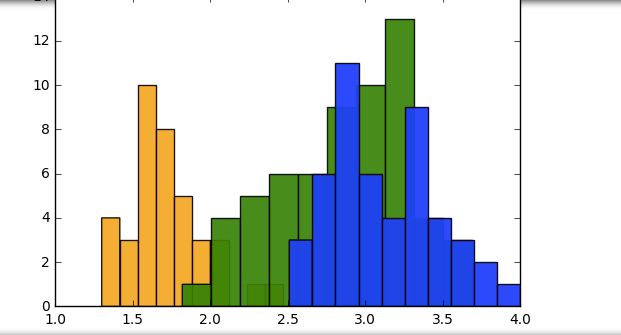
Variable 9:



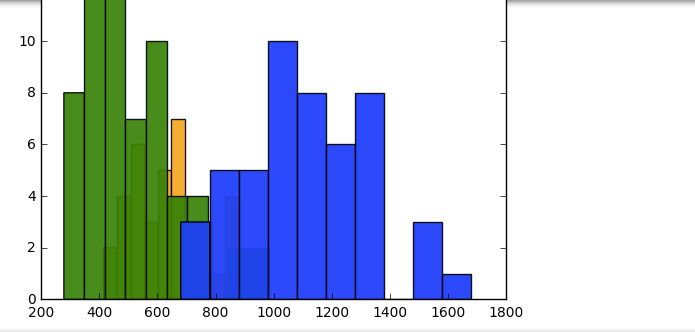
Variable 10:



Variable 11:

Variable 12:

Variable 13:

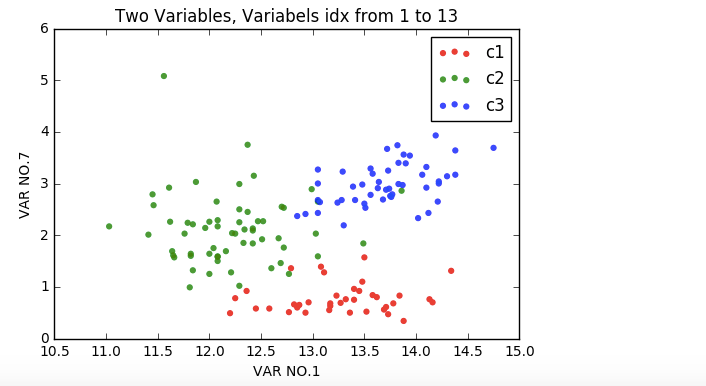


We can see the attributes of every class are concentrated.

3.4

H(x) is positive semidefinite for *all* x∈A, then f(x) is convex and has a strict *global* minimum at any

x for which f′(x)=0 and H(x) is positive *definite*.

3.5

We can guess the decision boundary, but still there may be some points can’t guess correctly.

3.6

Set accuracy 0.9 as stopping criterion. Because it can’t converge better like 3.1, hence, sometimes more attributes for classification is necessary.

