

No free lunch: Averaged over all problems, no machine algorithm is expected to perform better than any other algorithm.

Assumption when proof: true hypothesis f is uniformly distributed

Model selection: train-test split; train-val-test split; k-fold CV

Maximum Likelihood Estimation of 𝒘 under the Gaussian noise assumption is equivalent to linear regression with the least square loss

Computational complexity!!

Curse of Dim: neighbors are not local; data are very sparse. Needed number of samples grows exp with d.

e.g. unit ball: most data closer to the edge

Lec-2 LM

Lec-1 Intro

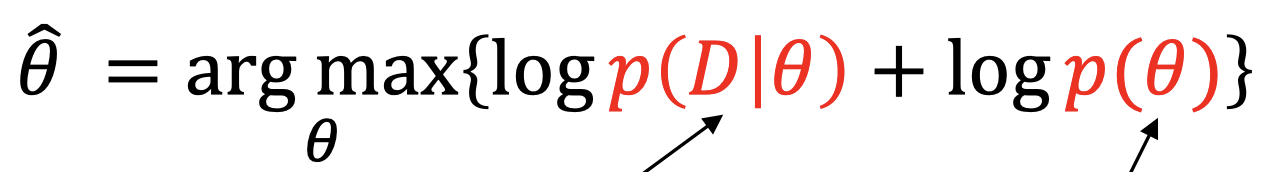
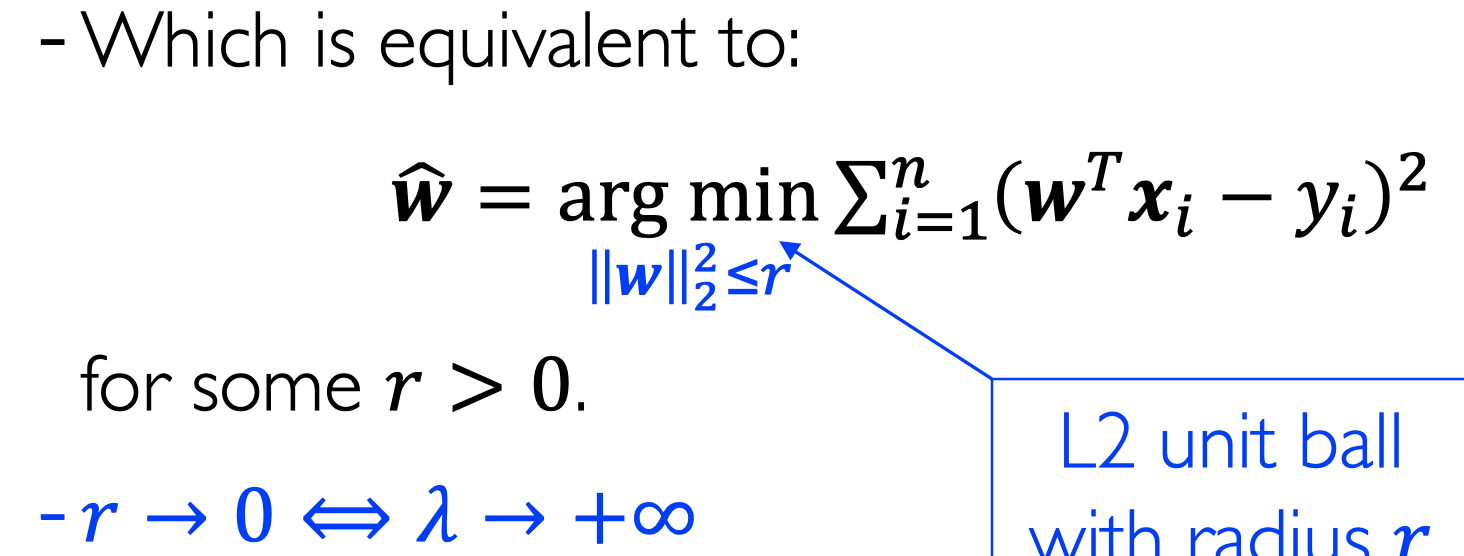
Nearest centroid classifier: computed per-class centroid

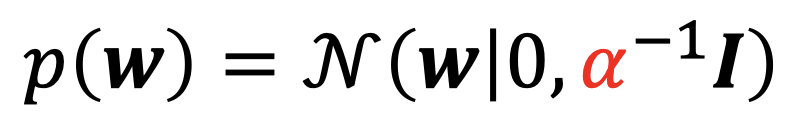
Weighted KNN

KNN: k less，train error smaller

Bayes optimal classifier:

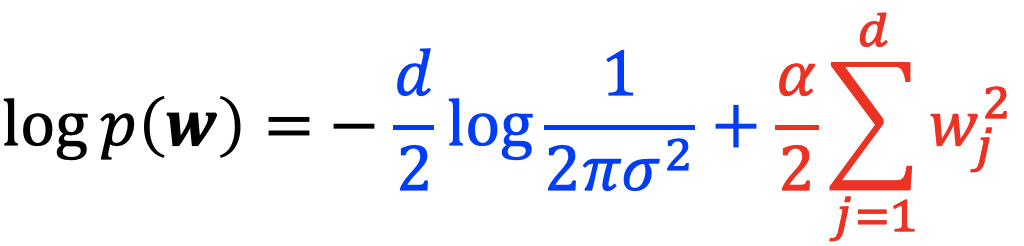
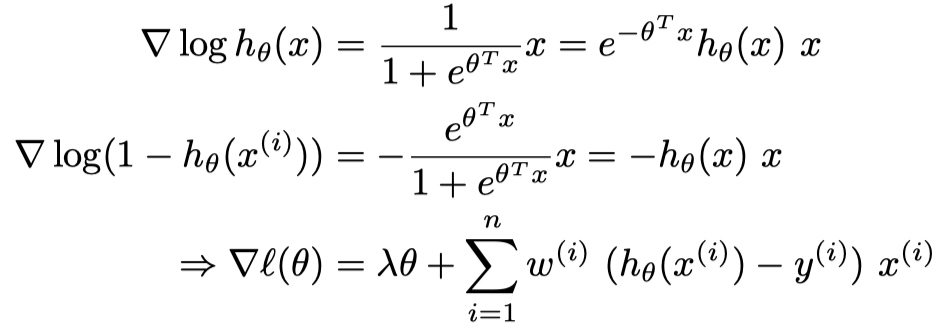
Learning model = hypothesis space (e.g. many kernel SVM) + learning algorithm (e.g. SGD, Adam)

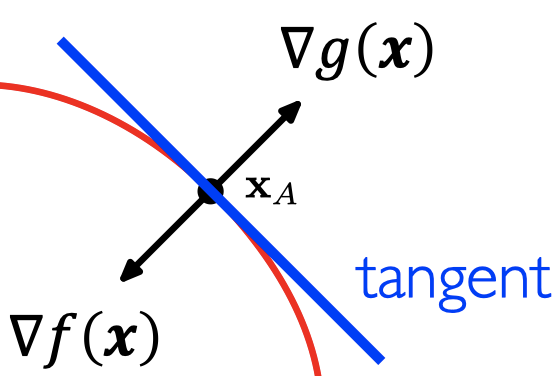
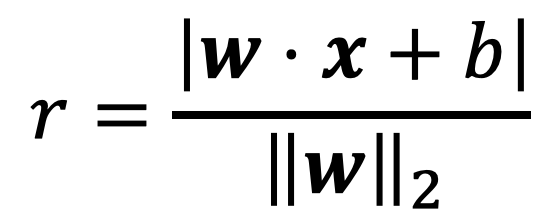
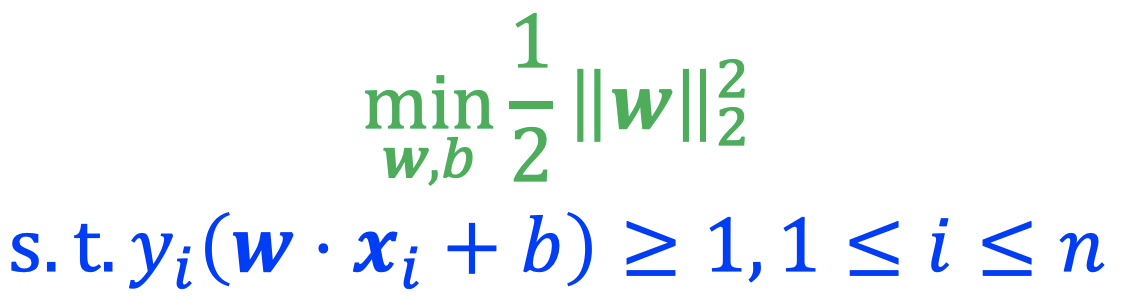
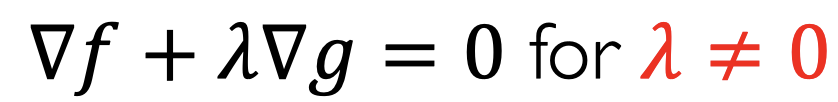
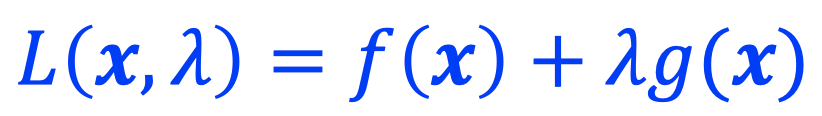
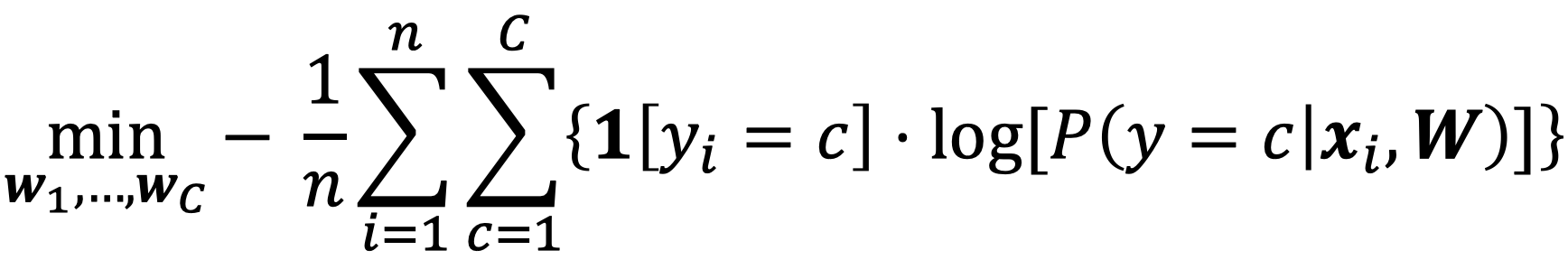
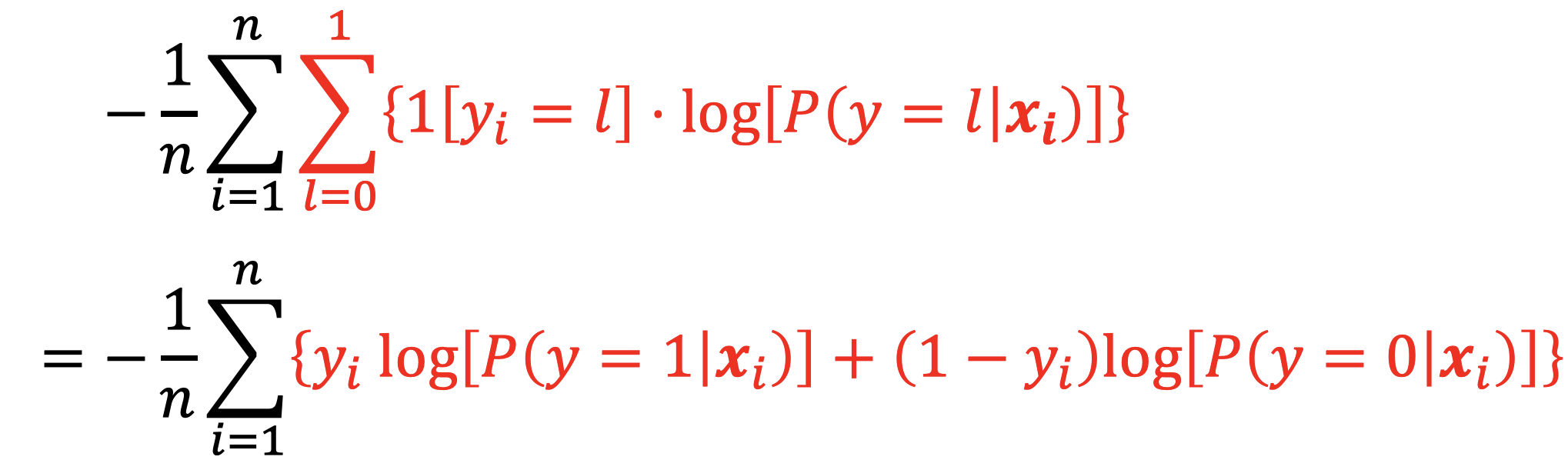
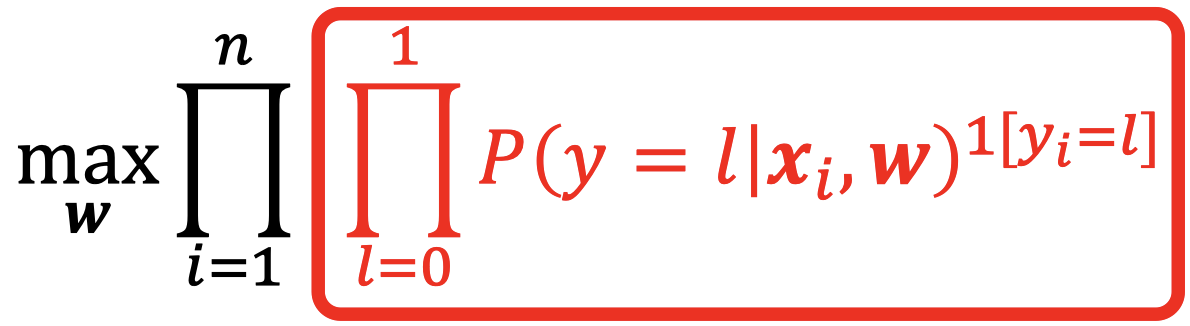
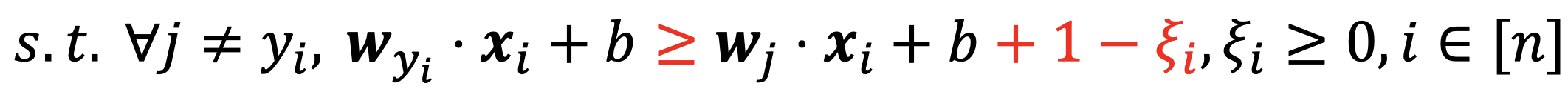
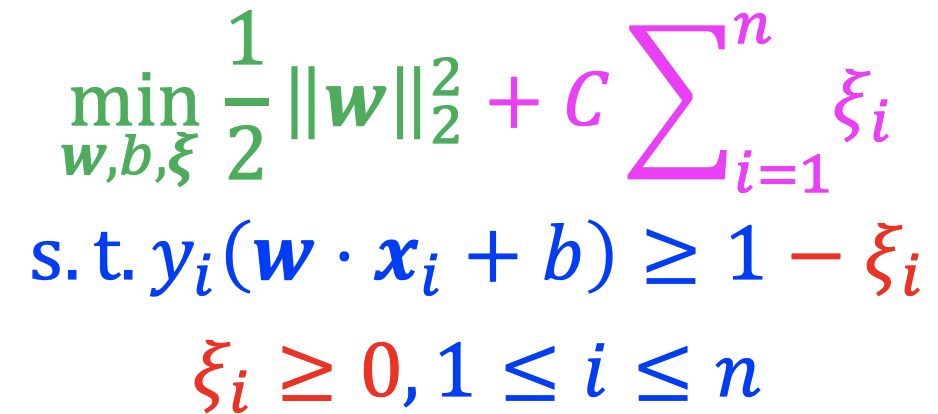
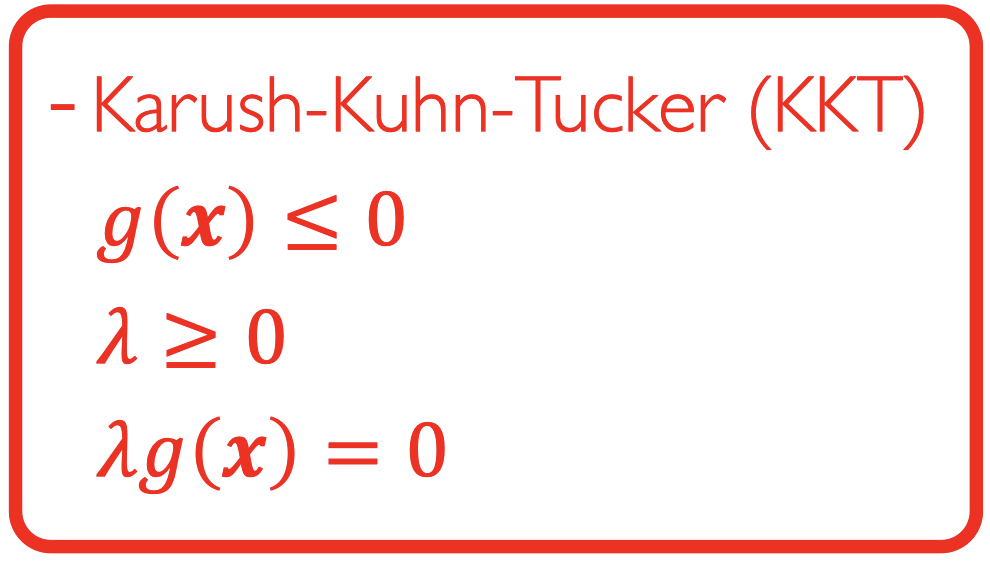
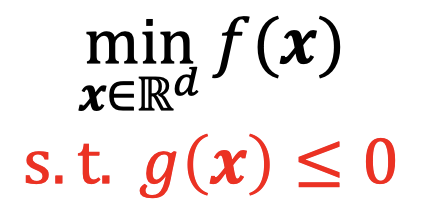
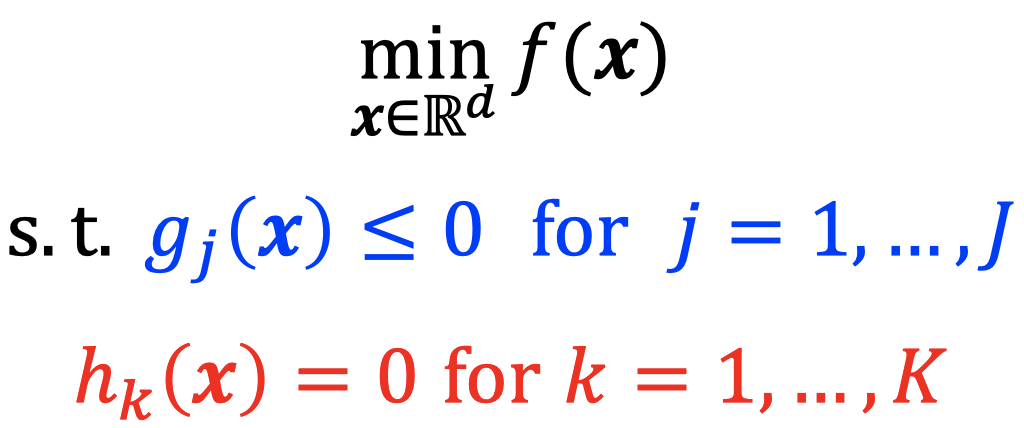
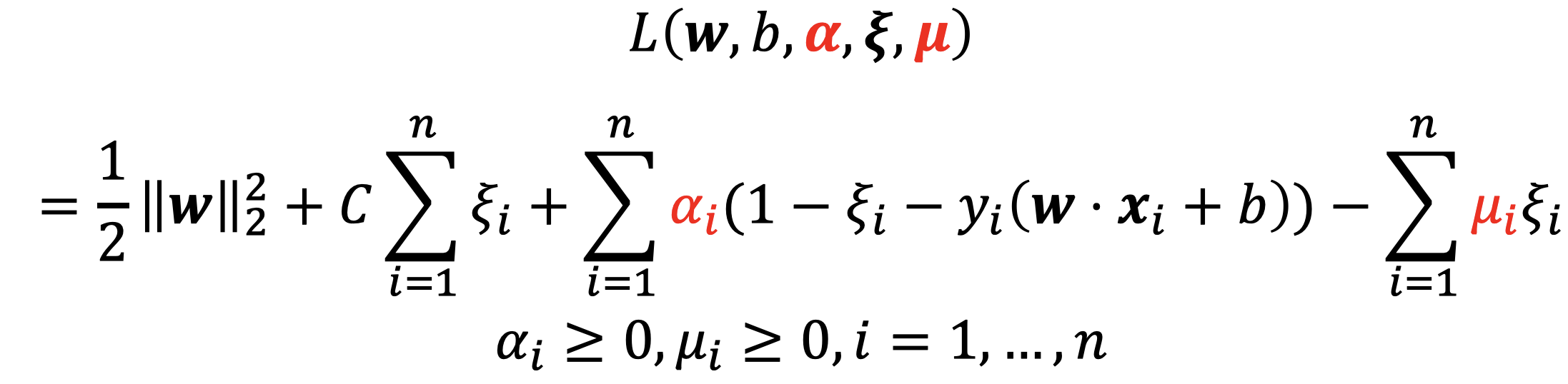
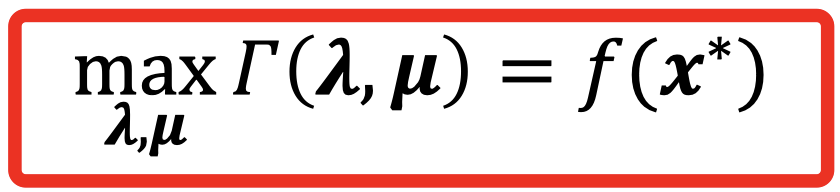
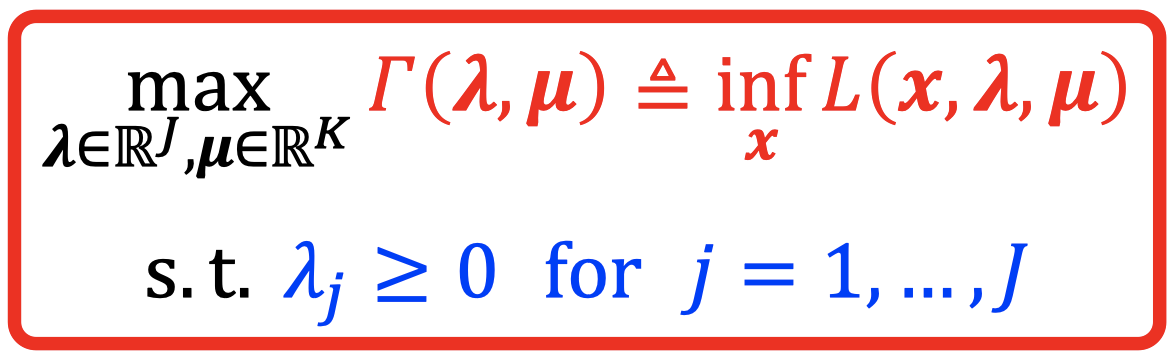
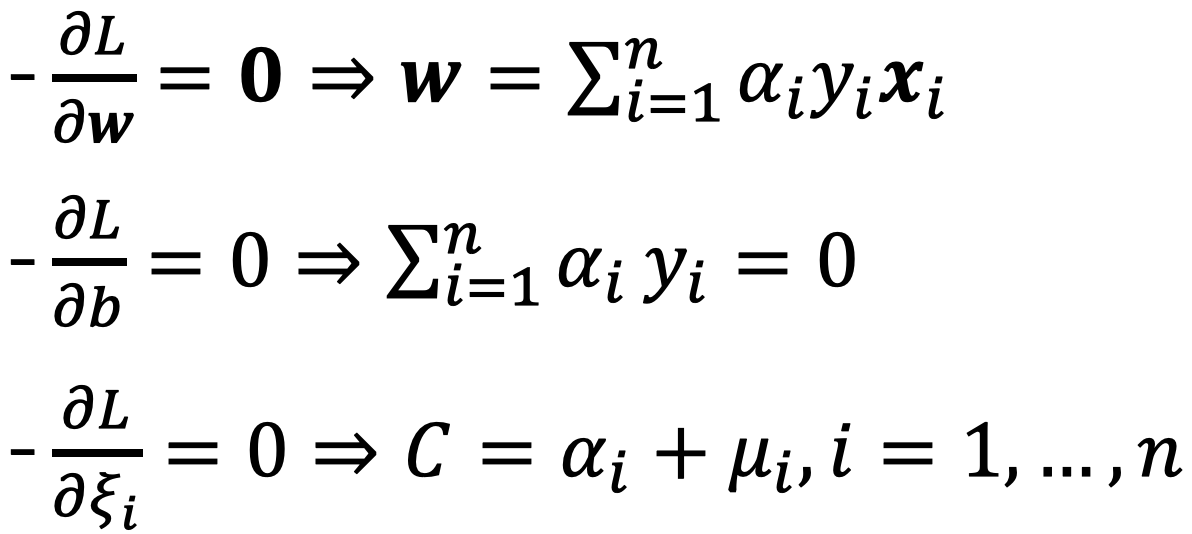
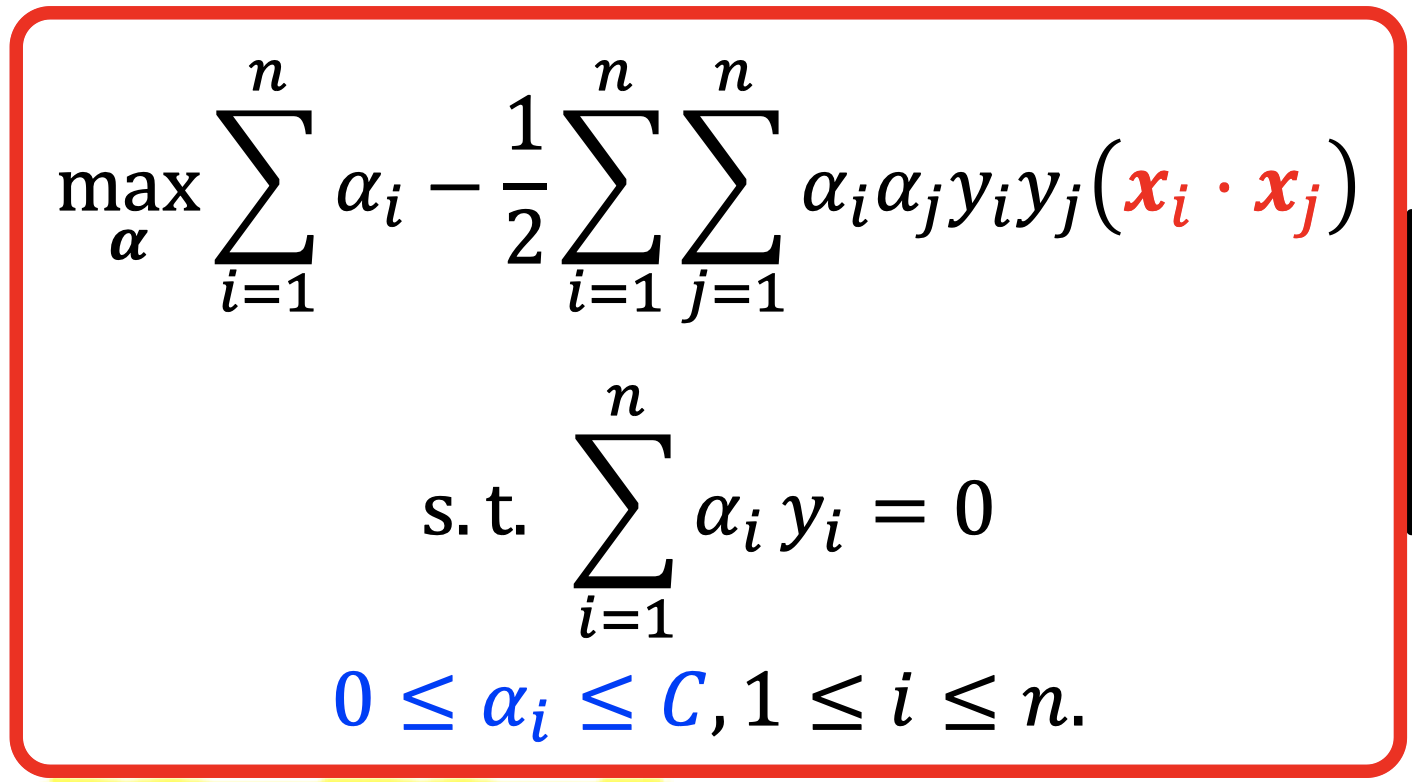
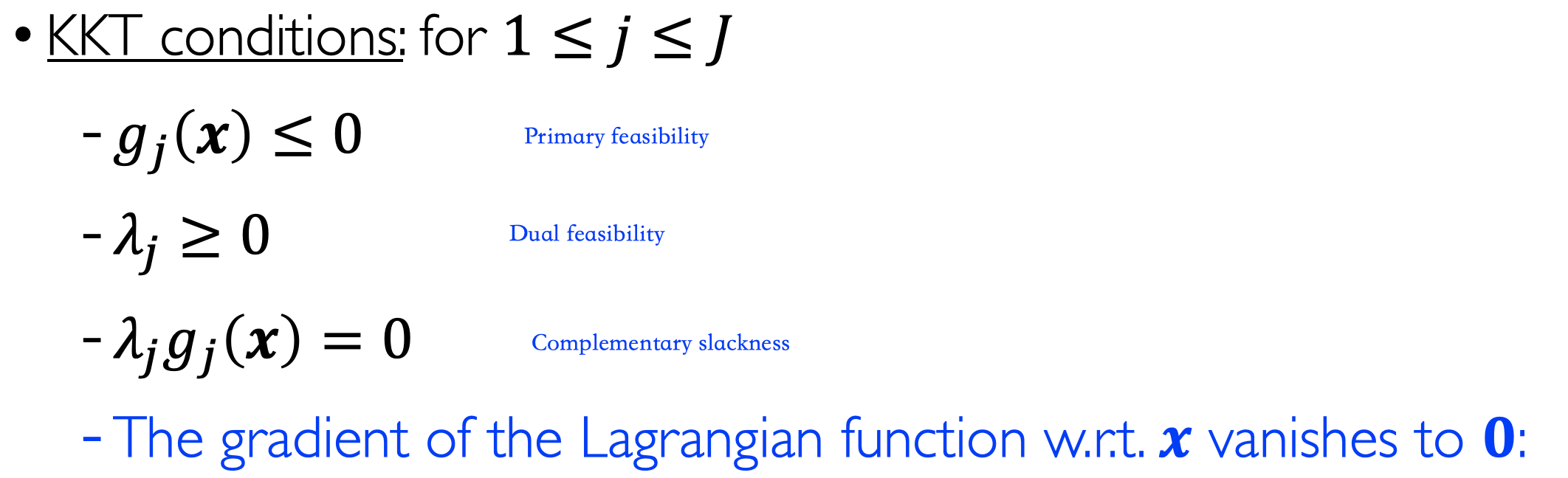
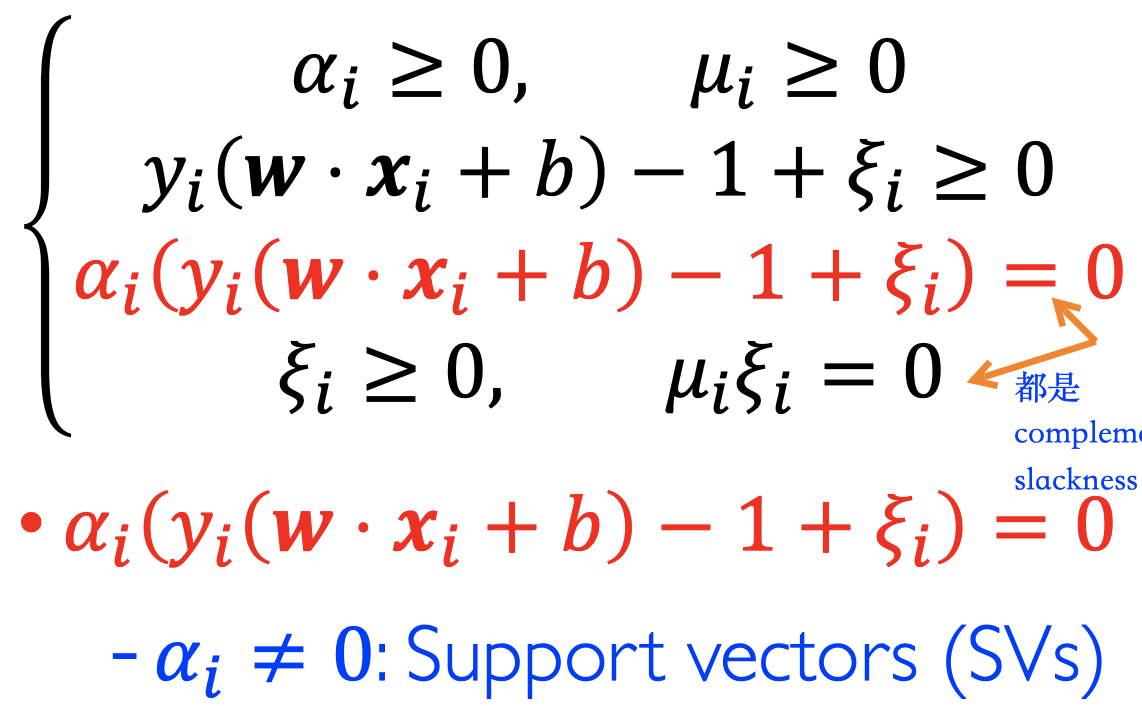
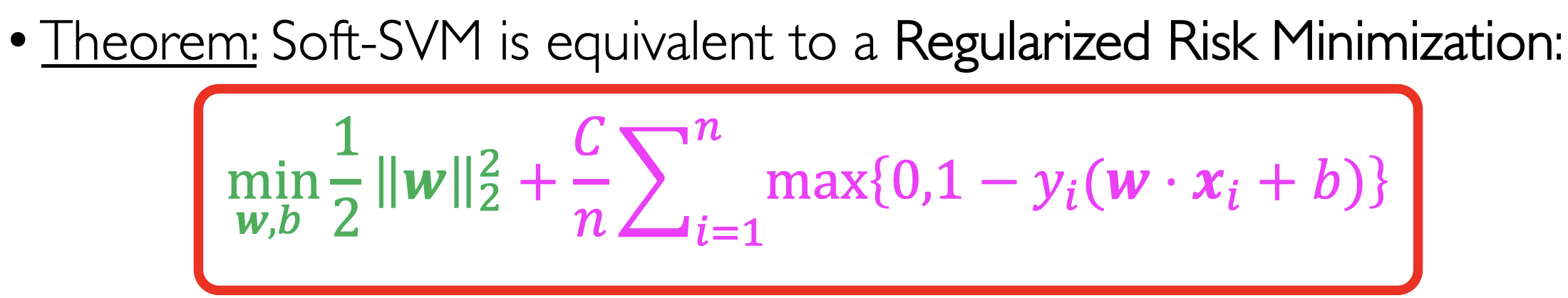
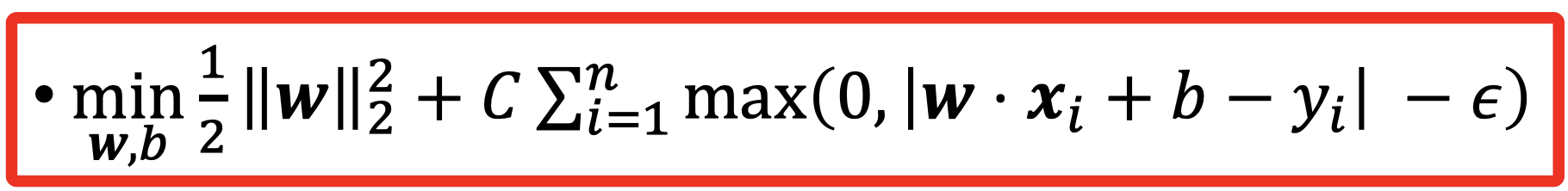
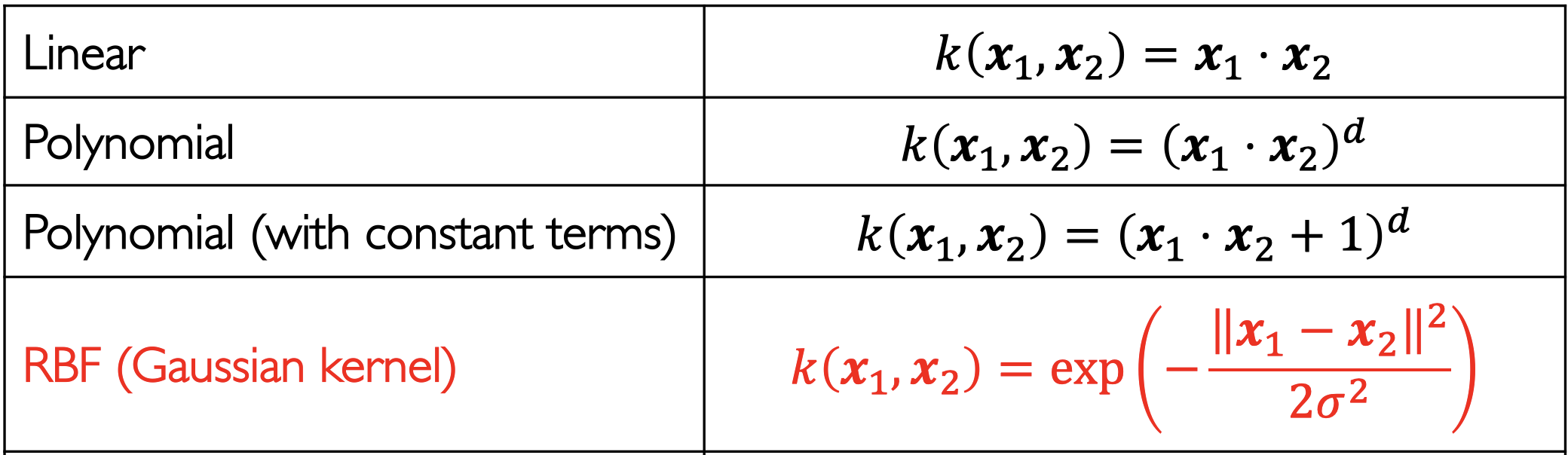
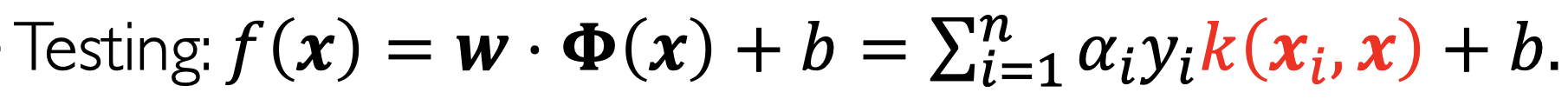
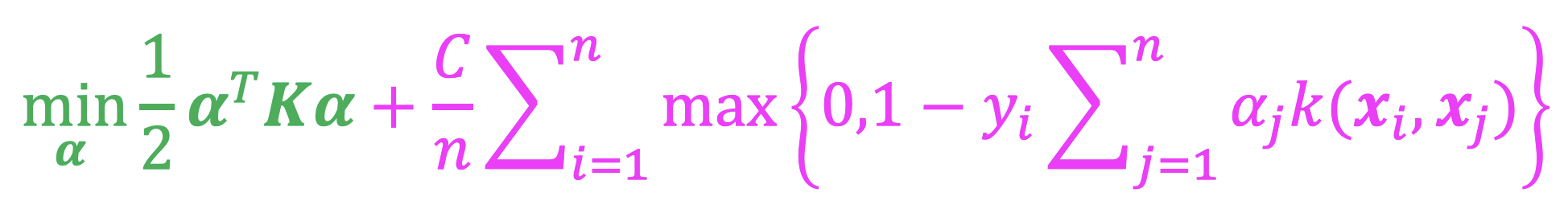
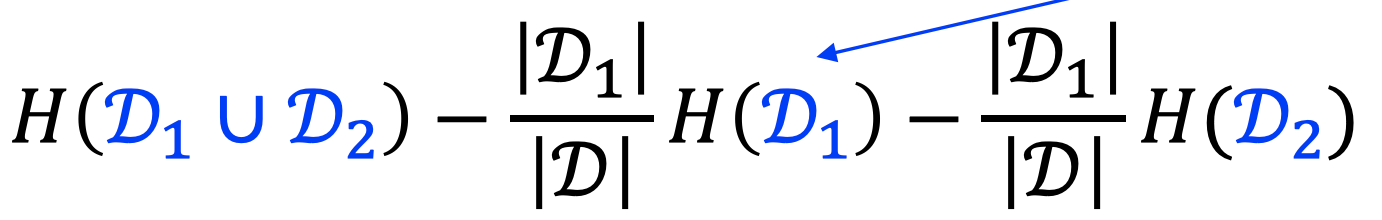
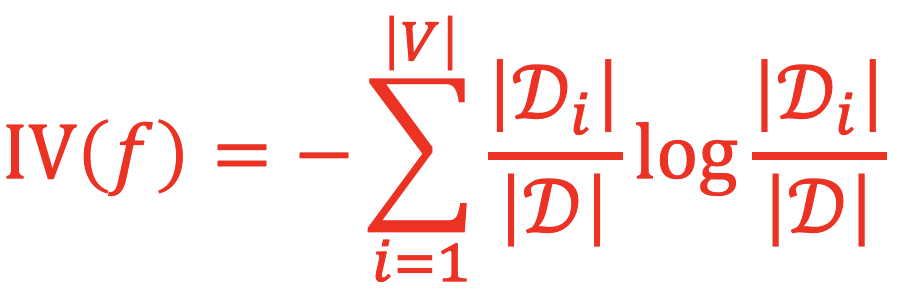
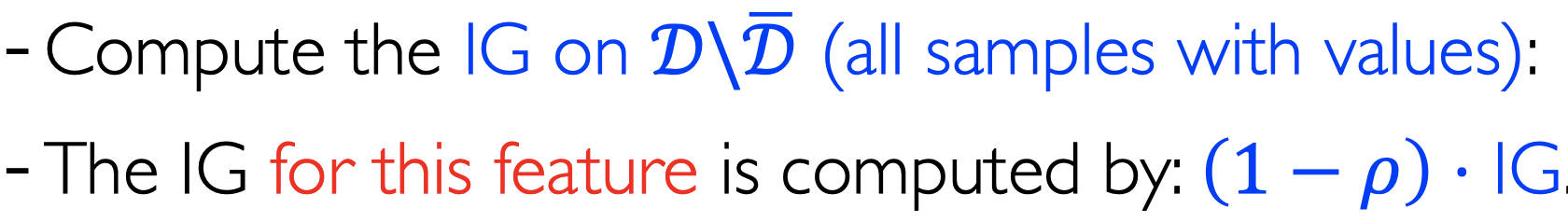
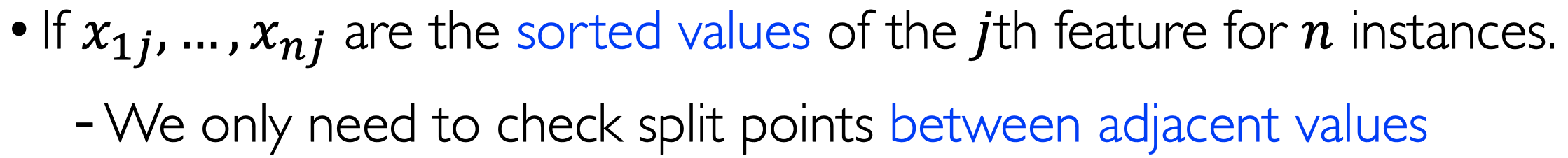
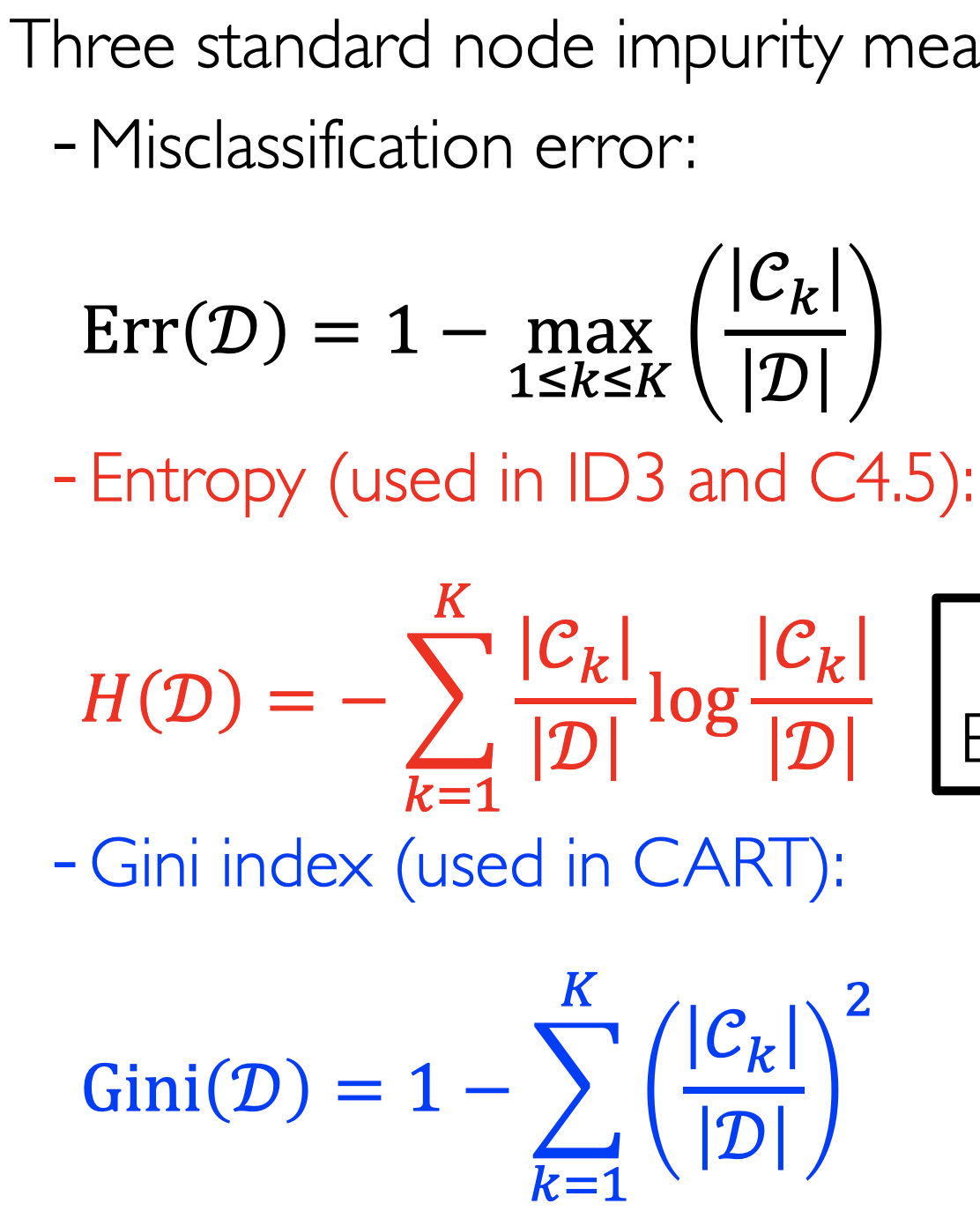




Ridge reg: L2;

Lasso reg: L1



During test: supplement missing values with the most common one

Information Gain is highly biased to multivalued features

ID3 to C4.5: IG Rate, attribute with costs, missing value, continuous value, pruning

Information Gain

Lec-5 tree

Representer Theorem：

SVR：

Convex:

g(x)<0 || g(x)=0

Lec-3 SVM

Multi-class