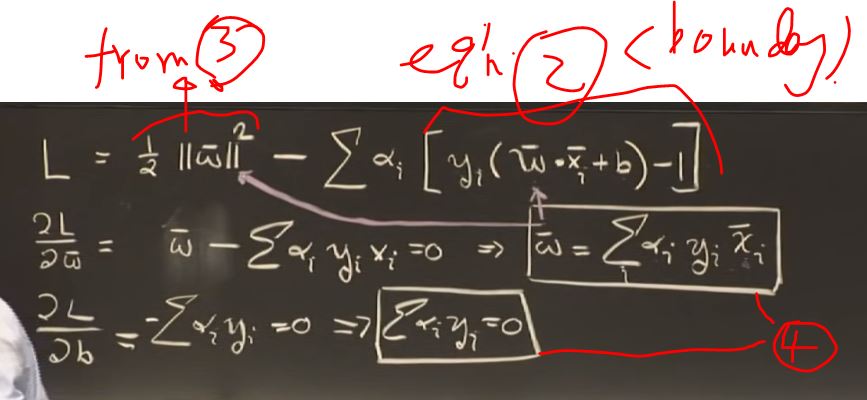


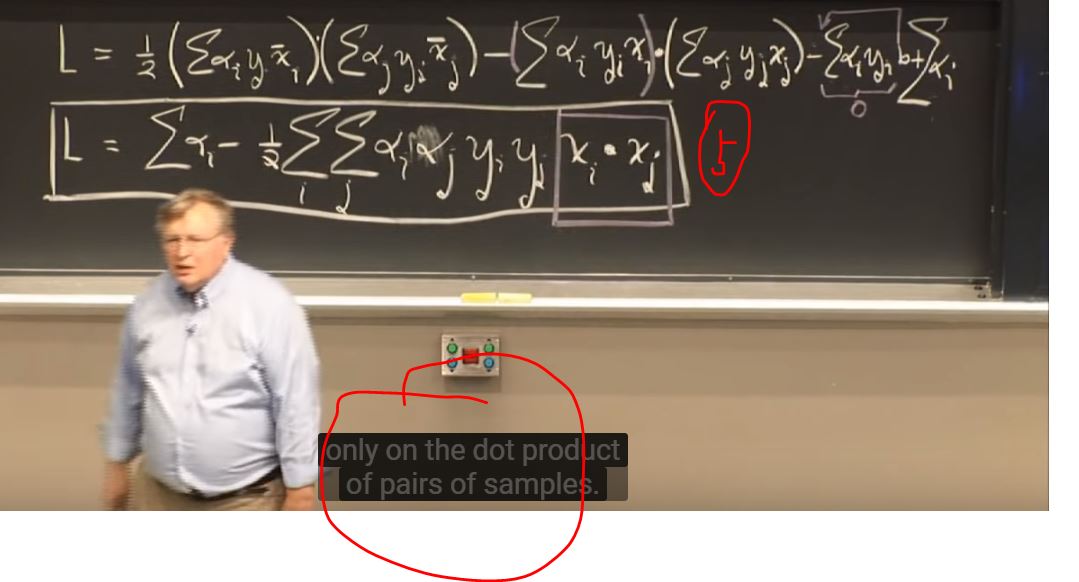
To find the extreme value of a function (e.g. function 3) with constraints/boundary condition applied (e.g. eq’n 2), we need to use Lagrange multipliers (alpha as in function below): -- principle – the extreme of the function 3 **with constraints (eq’n 2) applied** happens at the extreme of the function L (Lagrange function) as below.

So we only need to find the extreme of the function L as below without worrying about contraints (eq’n 2) cos it was already included in the function. So we just need to let the partial derivative equal to 0 as below. (vector w and vector b are the two variables)

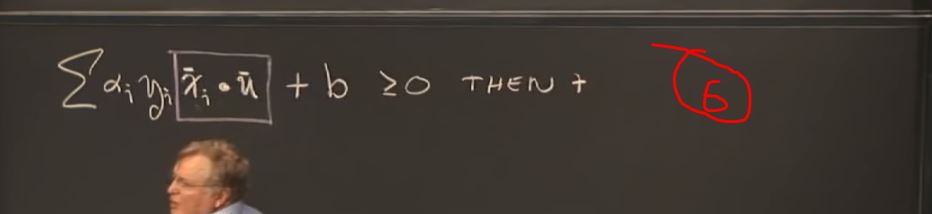


Sub eq’n 4 back to Lagrange equation, then we have eq’n 5 as below. 🡪 take home message:

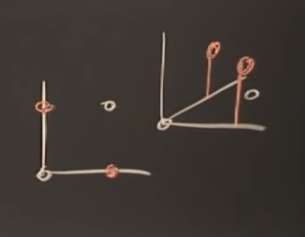
The extreme of the Lagrange function below ONLY depends on the dot product of the pairs of the samples (e.g. vector x i and j)

Sub Eq’n 4 back to Eq’n 1(decision rule), we can easily get that 🡪 take home message:

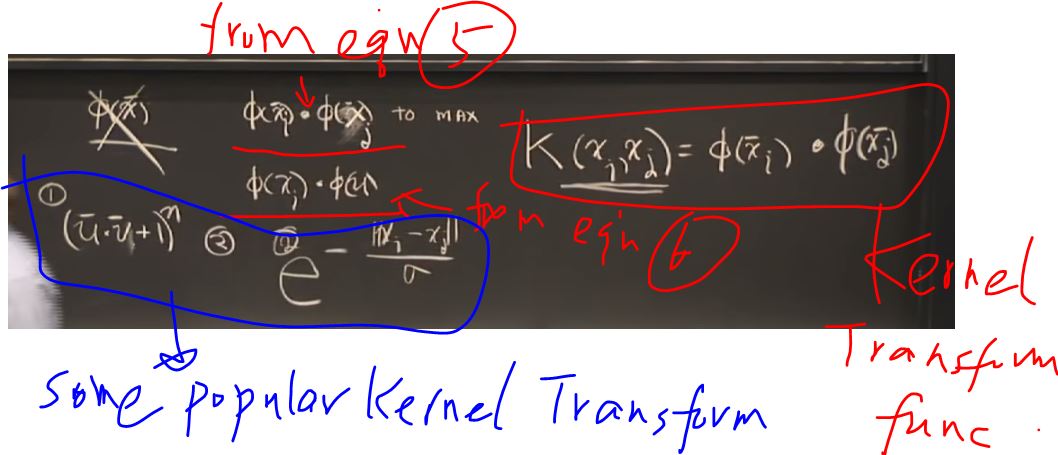
Decision rule also depends on the dot product of the sample vectors (e.g. vector x) and unknown vector (e.g. vector u).



But in some cases, like LHS below, we cannot separate two classes at existing dimension. We need a Kernel transformation to make them separable like RHS below. (e.g. create another dimension).



From Eq’n 5 and 6, we know to achieve largest width, **both decision rules and Lagrange depend only on dot product of the vectors.** So when we cannot make separation using existing dimension (i.e. xi\*xj or xi\*u), we switch to another dimension (i.e. kernel transformation of the two vectors as below). We can see some popular Kernel Transforms here as marked blue.



The rest of the proof can be handed over to mathematical experts. And we have some well-building tools for SVM to effectively separate different classes (e.g. as below)

