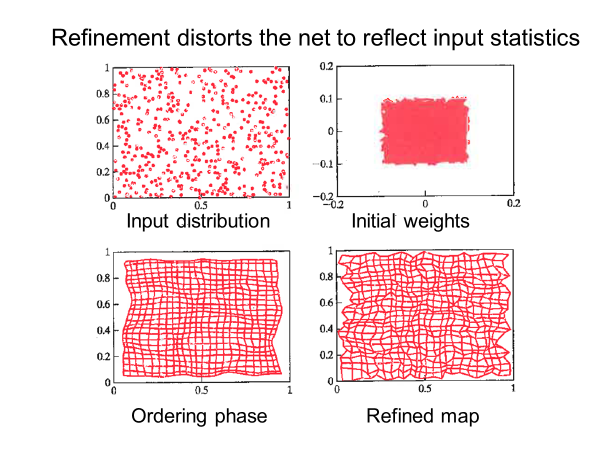
# Self-Organizing Maps (SOM)

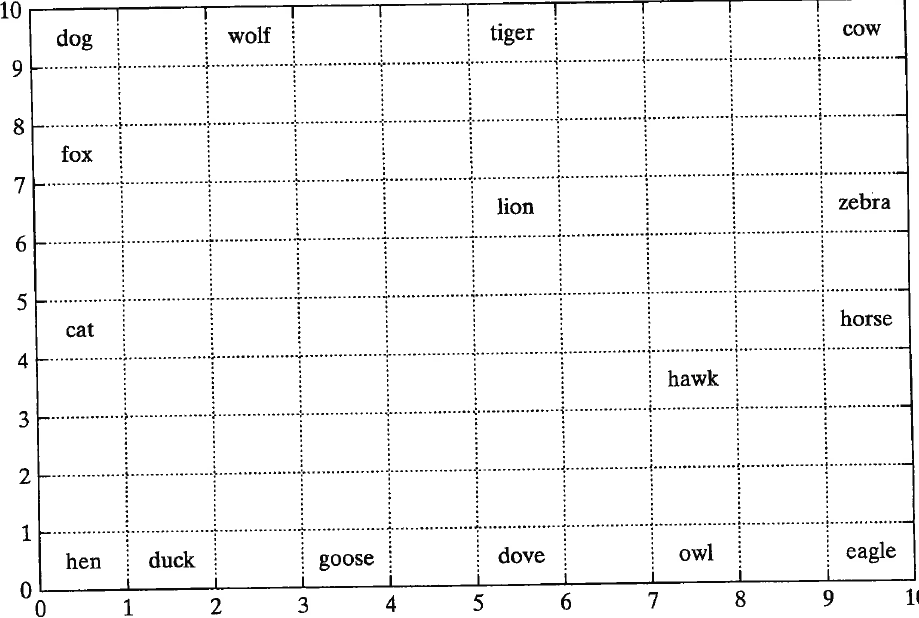
The 3 aspects of training in SOM algorithms:

1. Output nodes compete for activation based on a discriminant function.
2. Winning node, with largest value of discriminant, becomes the center of a cooperative neighborhood.
3. Neighborhood adapts to an input pattern because cooperation increases the susceptible to large values of the discriminant function.

|  |  |
| --- | --- |
| Elastic net over input space.  Each node is associated with a compressed representation of input space.  Phase of SOM convergence illustrated:  Ordering phase |  |

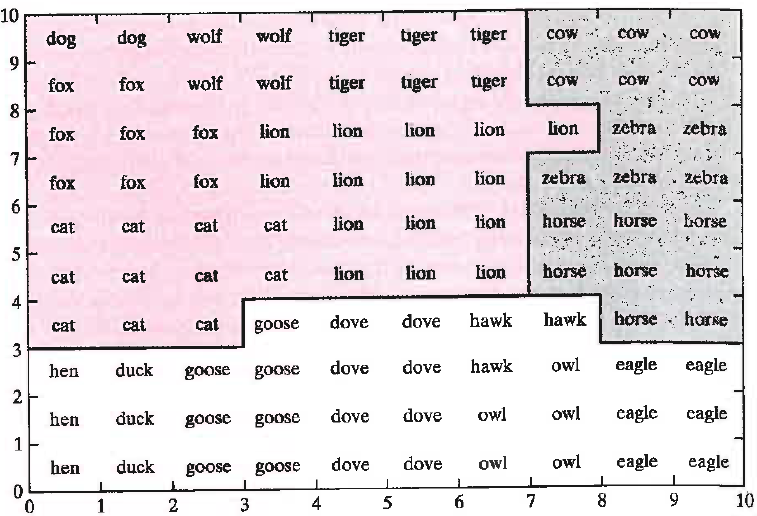


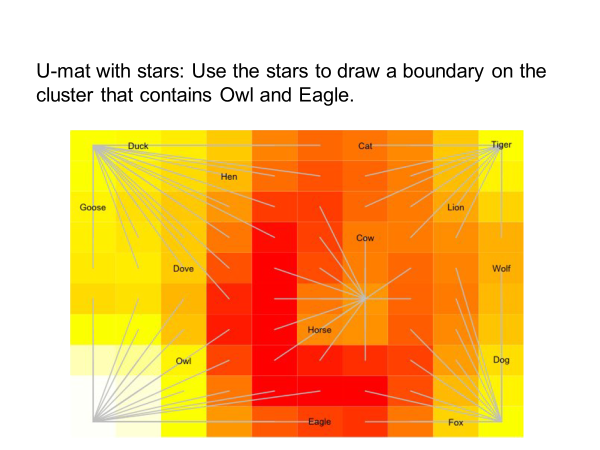
**Contextual map:** Labels output nodes based on responses of the lattice to “test patterns”; labeled instances, not used in training. Labeled node has strongest response to the indicated animal type. Topological ordering of data.



**Semantic map**: Lattice sites label by animal type inducing the strongest response at that node.

Similar types of data are clustered.





# Support Vector Machines (SVM)

SVM is an example of constrained optimization.

Simplex method can be used when both objective function and constraints are linear.

Why can’t the simplex method be used on SVMs?

In SVM, we find the optimal hyperplane (aka decision boundary) by minimizing a quadratic objective function subject to linear constraints.

How do we know that the solution to the SVM optimization problem is unique?

In the SVM optimization problem, the quadratic objective function is convex and bounded from below.

## Active set in the quadratic programming problem

The active set determines which constraints will influence the final result of optimization.

What is the active set in the SVM quadratic programming problem?

In the SVM quadratic programming problem, support vectors define the active set: attribute vectors with constraints that have non-zero Lagrange multipliers.

## Solving constrained optimization by Lagrange multipliers

**NOT EXPECTED TO BE ON QUIZ**

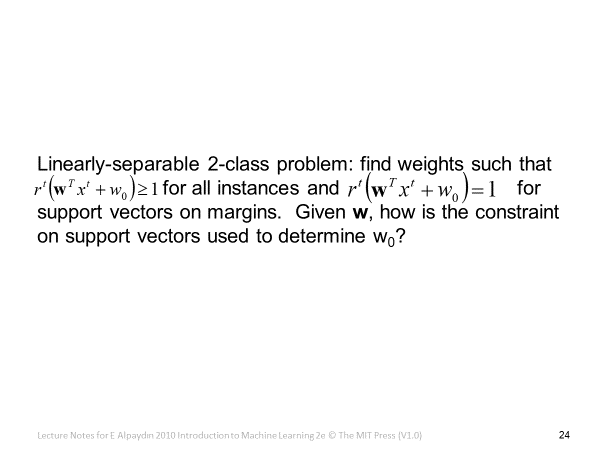
|  |  |
| --- | --- |
|  | Bishop Fig E |

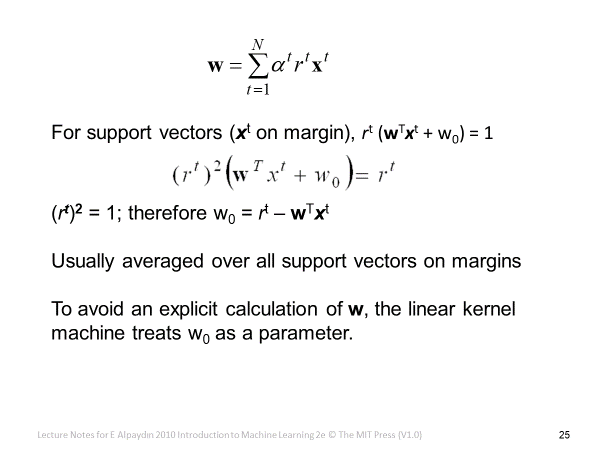
In the SVM constrained optimization, we maximize a “dual”; a Lagrangian with primal variables replaced by Lagrange multipliers.

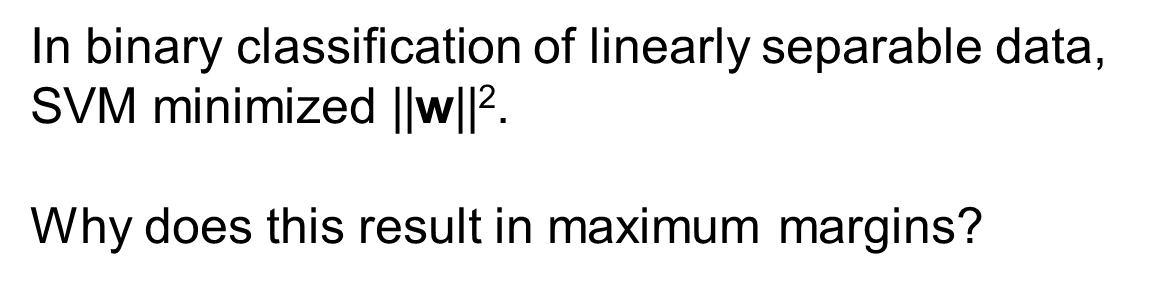
What calculus allows us how to make these replacements?

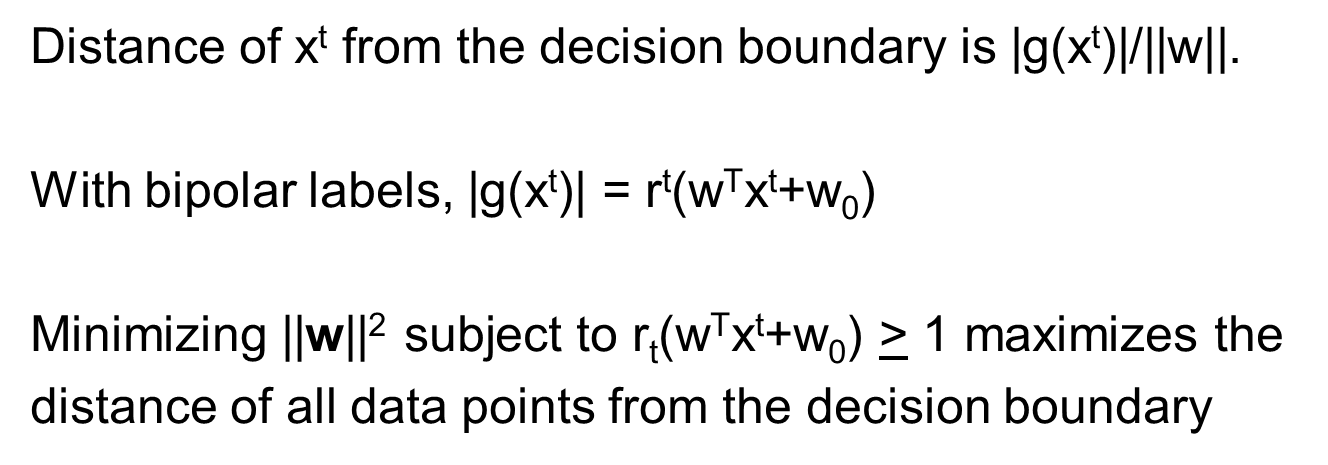
In constrained optimization with a Lagrangian, we set the partial derivatives of L with respect to primal variables and Lagrange multipliers to zero to find the stationary point.

This gives equations that allow primal variables to be written as a function of Lagrange multipliers.







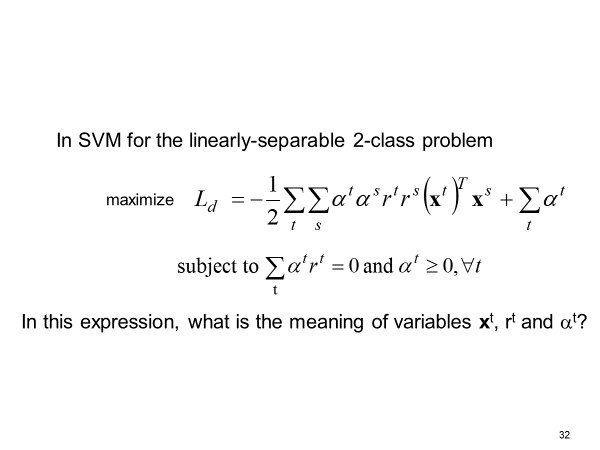


In constrained optimization, it is often possible to convert the primal problem (i.e. the original form of the optimization problem) to a dual form.

In general, solution of the dual problem provides only an upper bound on the solution of the primal problem (called duality gap). Why does maximizing the dual in SVM give an exact solution?

In SVM we always have a convex optimization problem with linear constraints.

For this type of constrained optimization, the duality gap is zero.

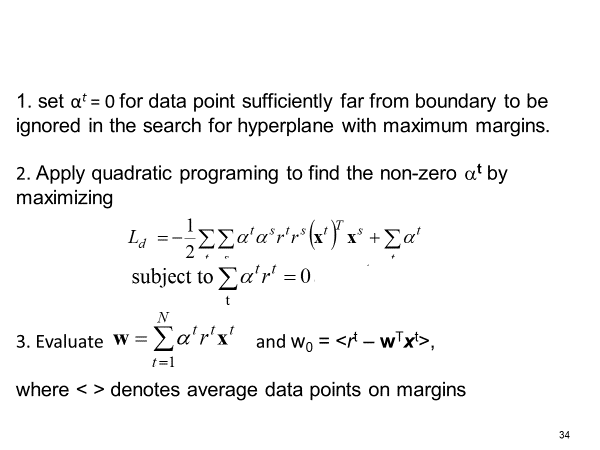


: Attribute vector

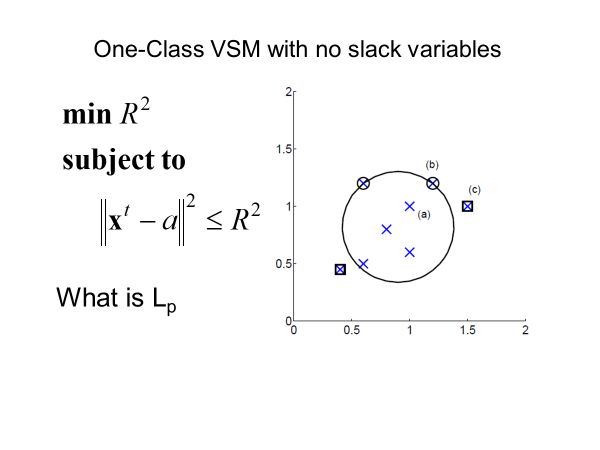
: Label on that attribute vector

: Lagrange multiplier that is the result of the constraint on that attribute vector

In SVM for the linearly-separable 2-class problem, what are the 3 steps to find the decision boundary with maximum margins?



## One-Class SVM with no slack variables

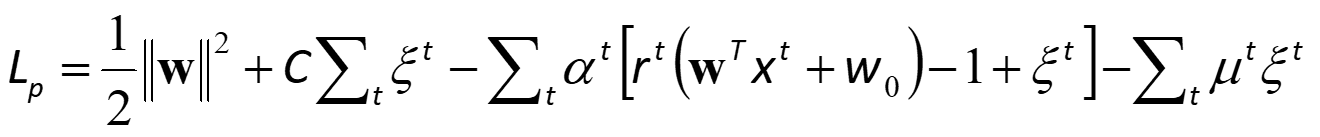




What is the derivative of with respect to R?

What constraint on Lagrange multipliers results from setting this derivative equal to zero?

## C-SVM: binary classification of non-linearly separable data



What are the primal variables?

, , and

What are the dual variables?

and are Lagrange multipliers.

What role does constant C play?

C is the regularization parameter for the penalty on soft error.

How does increasing C likely effect the width of margins?

Increasing C penalizes soft error more likely decreasing the width of margins with fewer points in the margins.

## n-SVM: alternative to C-SVM



What are the primal variables?

, , , and

What are the dual variables?

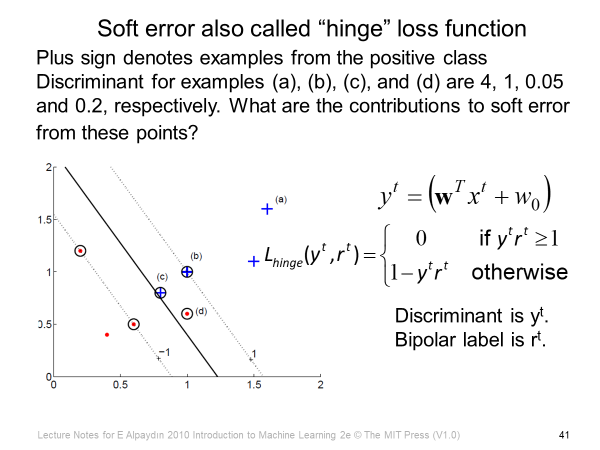
, , and

What is the meaning of regularization parameter ?

is the upper bound on the fraction of instances in margin.

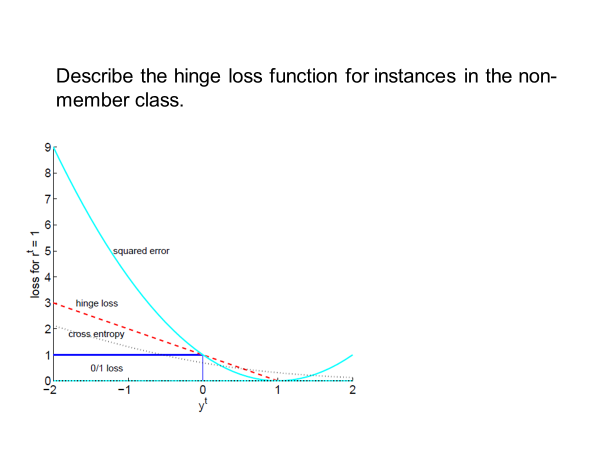
How does increasing likely effect the width of margins?

Increasing likely increases width of margins with more points in margins.



is +1 for members and -1 for non-members.

1. so
2. so
3. so
4. so



Solution: hinge loss

The above chart shows the hinge loss function for instances in the member class, since