We’ve look to SVM using soft margin function, now we look into it via Loss minimization function

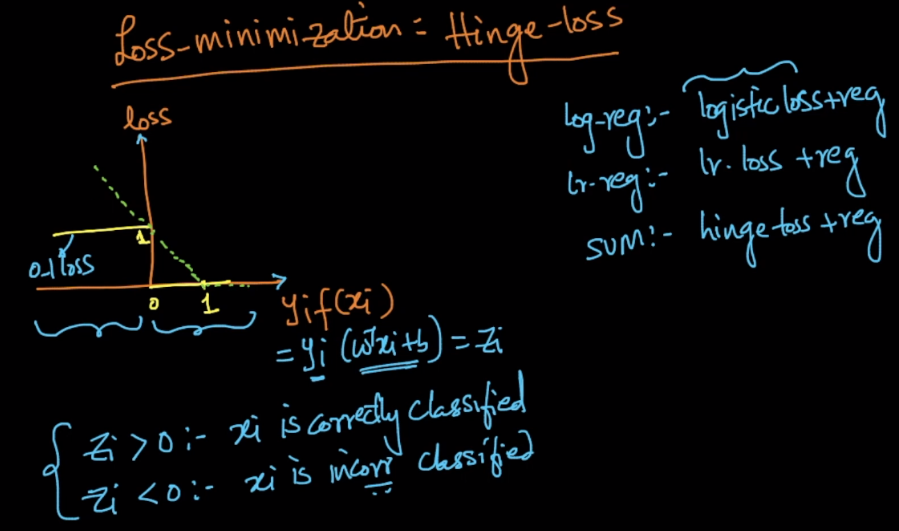
Why are we doing this, because as we seen in logistic and linear regression, just changing the loss function would give us the new model, so similarly in SVM there is a loss called **hinge loss**, so we just replace the loss with hinge loss and we get SVM.

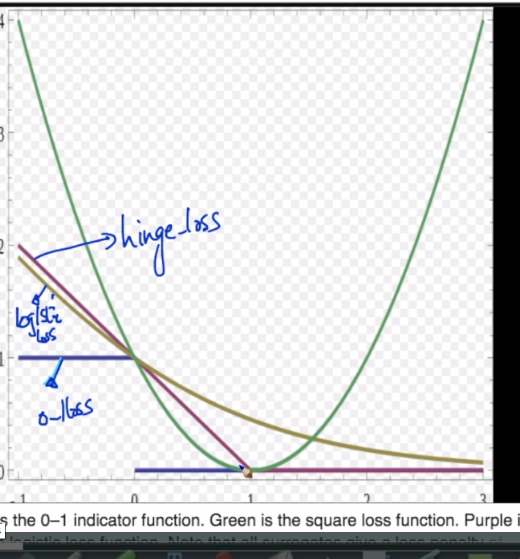
**What is hinge loss:**

Graph for hinge loss is shown below, on x axis we have z\_i as y\_i \* ((W\_T \* x\_i) + b), and on y we have hinge loss.

So as we can see that hinge loss become 0 when z\_i >= 1.

Hinge loss is not differentiable at z\_i = 1, because it’s not smooth here. Even though it’s continuous.





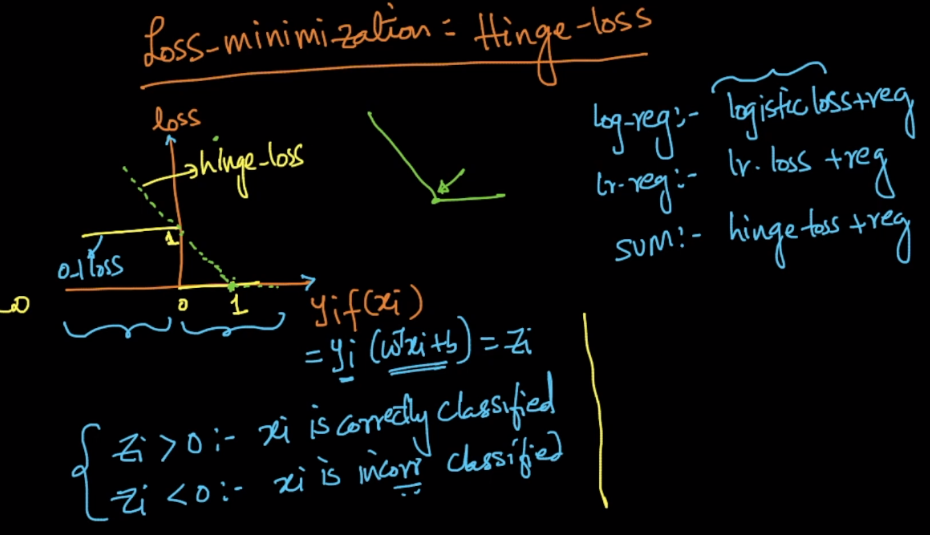
Now since we know that

Z\_i > 0 : if xi is correctly classified

And Z\_i < 0 : if xi is incorrectly classified.

Now what hinge loss says that if z\_i >= 1 loss become 0 and

For z\_i < 1, loss is 1 - z\_i.

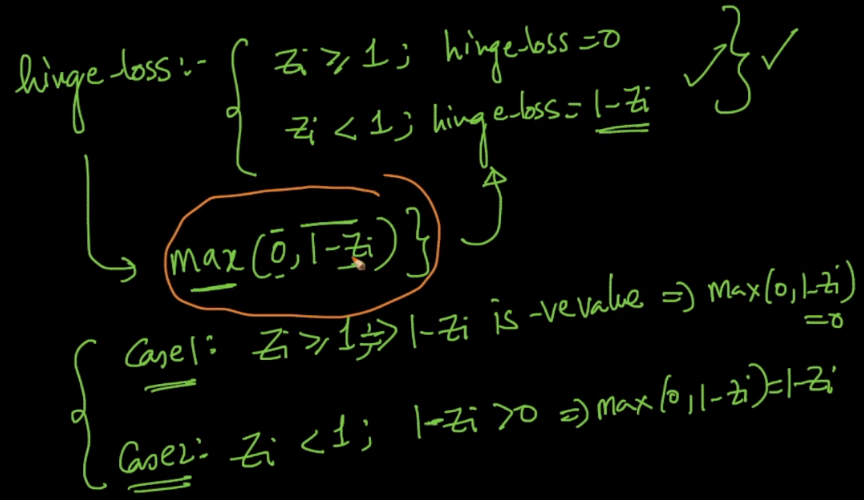


This hinge loss in one line can be written as **max(0, 1-z\_i)**

Why it can be written like this?

Case 1: when z\_i >= 1, that means it’s correctly classified and 1-z\_i will be –ve value and since 0 > -ve and loss for correctly classified should be 0, therefore it’s true for z\_i >= 1.

Case 2: when z\_i < 1, that means 1-z\_i > 0, therefore here there would be some loss.



**Geometric representation of hinge loss:**

Let’s say there is +ve point below pie-, so it’s distance from pie+ will be.

d\_j = 1 - y\_i \* ((W\_T \* x\_i) + b) = 1 – z\_i.

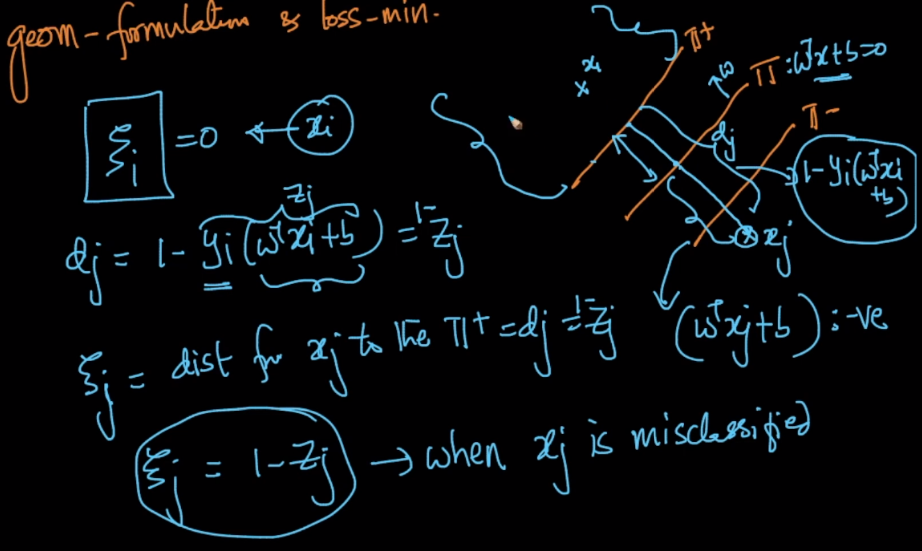
why for that +ve point lie below pie-, y\_i \* ((W\_T \* x\_i) + b) will be -1.5, and

therefore d\_j = 1 – (-1.5) =2.5.

And as we know zeta = d\_j.

So therefore we can say when point is misclassified that zeta = 1 – z\_i

**Here zeta is loss, as we can see that zeta is similar to loss we get in hinge loss for incorrectly classified points.**



**Let’s look Soft margin SVM and Loss minimization of SVM together.**

* As we can see that in margin SVM, right is loss and in Loss min, left is loss.

Here in margin SVM C is associated with loss but there is nothing in loss of loss min.

* Left of margin SVM is regularization , right loss min, is regularization.

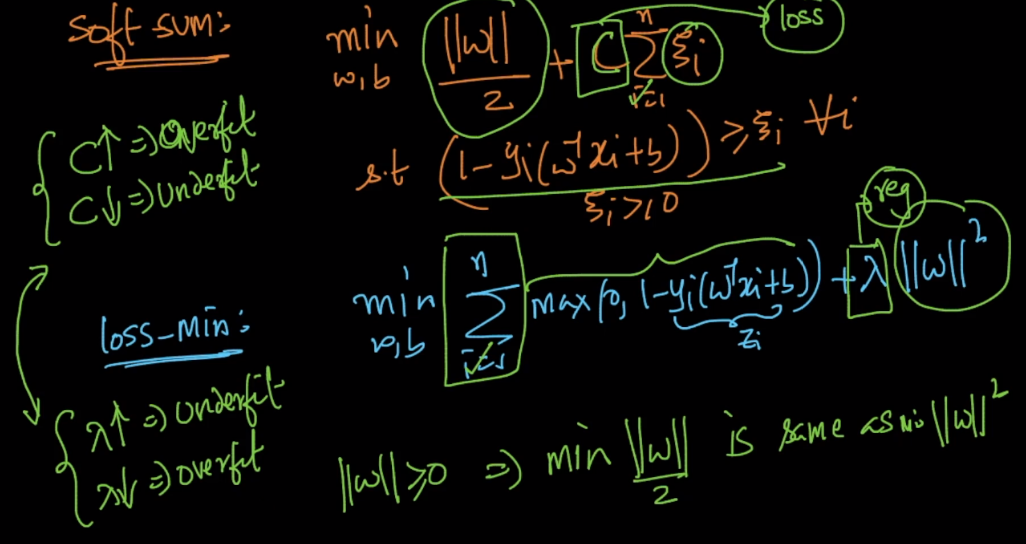
There is regularization term lambda associated in loss min, and there is nothing in loss min, regularization.

In margin SVM:

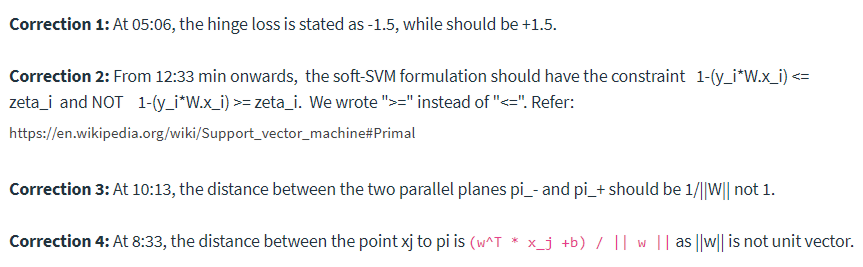
* C increases, become overfit
* C decreases, become underfit

In loss minimization

* Lambda increases, become underfit.
* Lambda decreases, become overfit.



Following are the corrections in images



<https://medium.com/@ashwanibhardwajcodevita16/from-zero-to-hero-in-depth-support-vector-machine-264931a1e135>