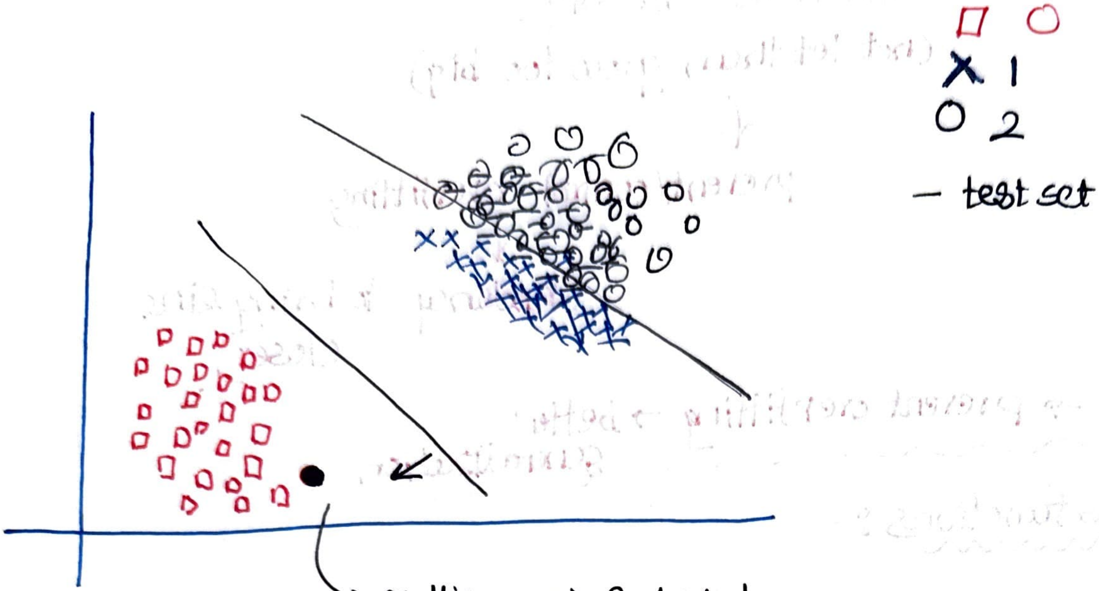
week 4

The effect of outliers



outlier mislabeled.

the line would more towards outlier

sometimes the outlier might cause significant change in separation.

w :- become big

(x_1, x_2) , $(x_1 + \varepsilon, x_2)$

little noise

if ε is relatively less compared to the data, we can expect prediction for both will be the same.

$z = w_1 x_1 + w_2 x_2 + w_0$

$\hat{z} = w_1(x_1+\varepsilon) + w_2 x_2 + w_0 = z + w_1 \varepsilon$

noise

allow w_1 to become too big. our neuron would give too much importance to ε (noise).

→ maybe very big weights are not good to have

Introduce

Regularization Bias :- forcing my weight to not be too big. (by adding some penalty)

Regularization • weights ω
Bias • Should not be big.
(not let them grow too big)
↓
prevent/control overfitting
↓
(tendency to bring line closer)

Regularization → prevent overfitting → better generalization

* Regularized loss functions :-

linear/logistic
→ penalize for having large weights.

$L(X, y, \omega) \longrightarrow$
regularized loss fn

$$L_{reg}(X, y, \omega) = L(x, y, \omega) + C^{-1} \| \omega \|_2$$

→ penalty → 2norm of weights
↘ penalty

l_2-regularization →
also called (ridge regression)

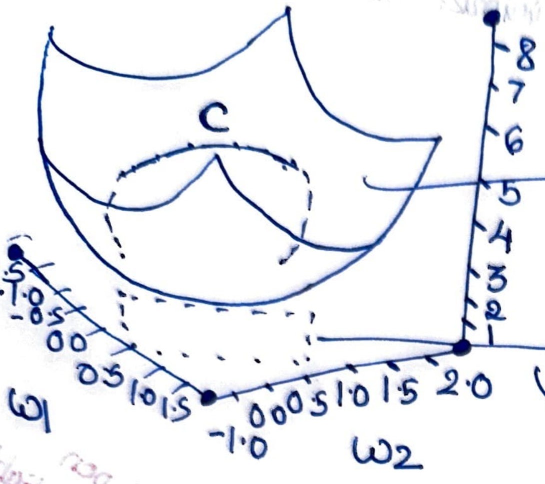
Hyper parameter → whenever C is very big value
we have no regularization
if C is moderate/small → Regularization
$C = 1$ (default) → sci-kit learn.

$$L_{reg}(X, y, \omega) = L(x, y, \omega) + C^{-1} \| \omega \|_1$$

→ I norm.
→ l_1 regularization.
(lasso regression)

$$L_{reg}(X, y, \omega) = L(x, y, \omega) + \underbrace{C_1^{-2} \| \omega \|_1 + C_2^{-1} \| \omega \|_2}_{\text{elastic net (combination of } l_1 \& l_2)}$$

* Geometric Understanding of Regularization:-

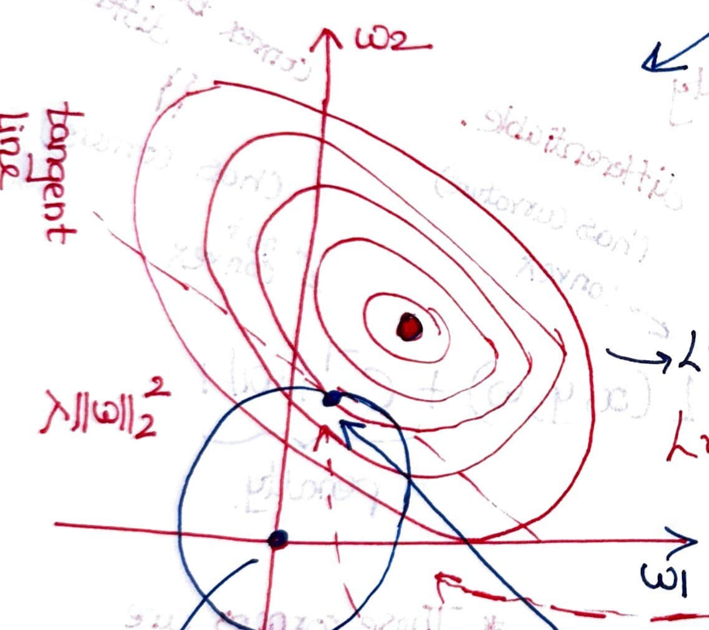


→ level sets of the loss function

$\mathcal{L}(\omega_1, \omega_2) = C$

→ Projett them in 2D space.

→ Contour Plots



→ $\mathcal{L}(X, y, \omega)$

$\mathcal{L}_{reg}(X, y, \omega) = \mathcal{L}(X, y, \omega) + \underbrace{C^{-1} \|\omega\|_2}$

also looks like another Loss fn (has countour plots)

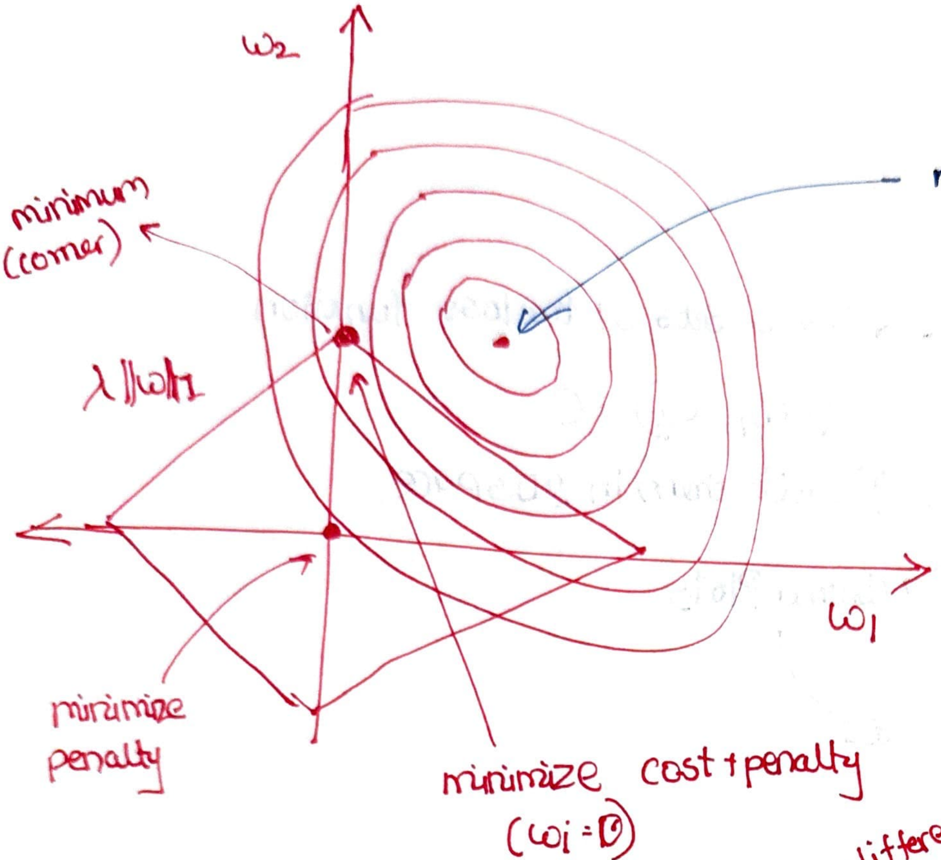
minimize penalty (minimum of penalty)

minimize cost + penalty

we find the balance between both. so not too much penalty/cost.

L_2 regularization.

L1 regularization



differentiable.
(has curvature)
convex

convex but non differentiable
{}
(has corners)
not convex

$\mathcal{L}(x, y, w)$

$$\mathcal{L}_{reg}(x, y, w) = \mathcal{L}(x, y, w) + \underbrace{C^{-1} \|w\|_1}_{\text{penalty.}}$$

* Corner of Square → Some coordinates will be 0.

(w2 was dropped)

* These corners are aligned in a way where they penetrate deeper into the contours of the usual loss function

lot of ws which are 0

L_2 is better (smooth)

l_1-regularization	// select attributes // small models	} good