Different Types of SVM model-Contd.

A. **Kernel SVM:** Kernel SVM works on the dual of the SVM model such that it is a Similarity check between two data points x_1 and x_2 :

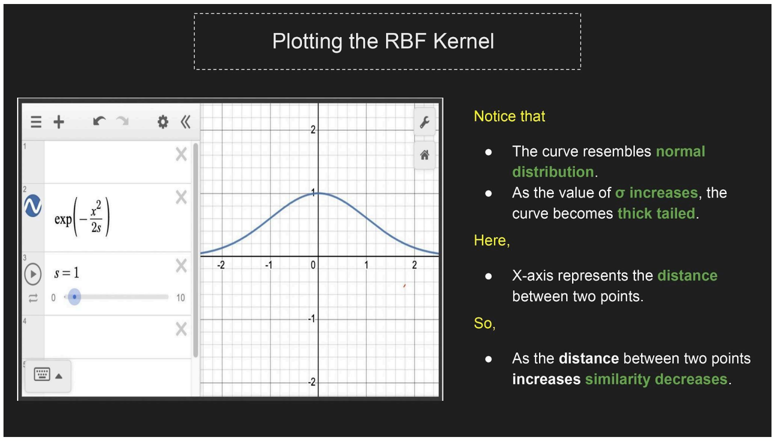
$Kernel\ Function\ =\ k(x_1, x_2)\ =\ (x_1^{\ t}x_2 + c)^m$, where m is the degree of Polynomial

Understanding Kernel Trick

- **Kernel Trick:** Kernel function projects a d-dim data to d'- dim data (where d'>>d) such that the data points are easily separable
- Can use kernelization with LogReg and Deep Learning models also.

Popular kernels

- Quadratic kernel: here m=2
    - $K_q(x_1, x_2)\ =\ (x_1^{\ t}x_2 + c)^2$
    - Now if C=1 , then on expanding $K_q(x_1, x_2)$,it becomes a 6 dimensional data.
- RBF(Radial Basis Function) / Gaussian kernel
    - Effective when not sure which degree of polynomial (m) to use
    - $K_{rbf}(x_1, x_2)\ =\ e^{\frac{-||x_1 - x_2||^2}{2\sigma^2}}$; where σ is a hyperparameter and $||x_1 - x_2||^2$ is euclidean distance



Plotting the RBF Kernel

Notice that

- The curve resembles **normal distribution**.
- As the value of **σ increases**, the curve becomes **thick tailed**.

Here,

- X-axis represents the **distance** between two points.

So,

- As the **distance** between two points **increases similarity decreases**.

RBF kernel SVM vs kNN

- RBF kernel SVM is similar to KNN geometrically but differs in runtime complexity: **O(#SV * d),** where
  - #SV: Number of support vectors;
  - d: number of dimensions

How is the Loss Function calculated for SVM?

The Loss function consists of two components:

- **Hinge Loss:** To have a minimum ζ , such that $y\ (w^T x\ \ + b)\ \geq\ 1 -\ \zeta$

$$\textbf{Hinge Loss: } \frac{1}{n}\sum_{i=1}^{n}\zeta_i$$

- Max Margin: To have a maximum Margin for generalizing on the data

$$Max\ Margin\text{: } \frac{||w||}{2}$$

Thus the total loss becomes:

$$\textbf{Loss: } min_{(w,b)}\frac{||w||}{2} + C\,\frac{1}{n}\sum_{i=1}^{n}\zeta_i$$

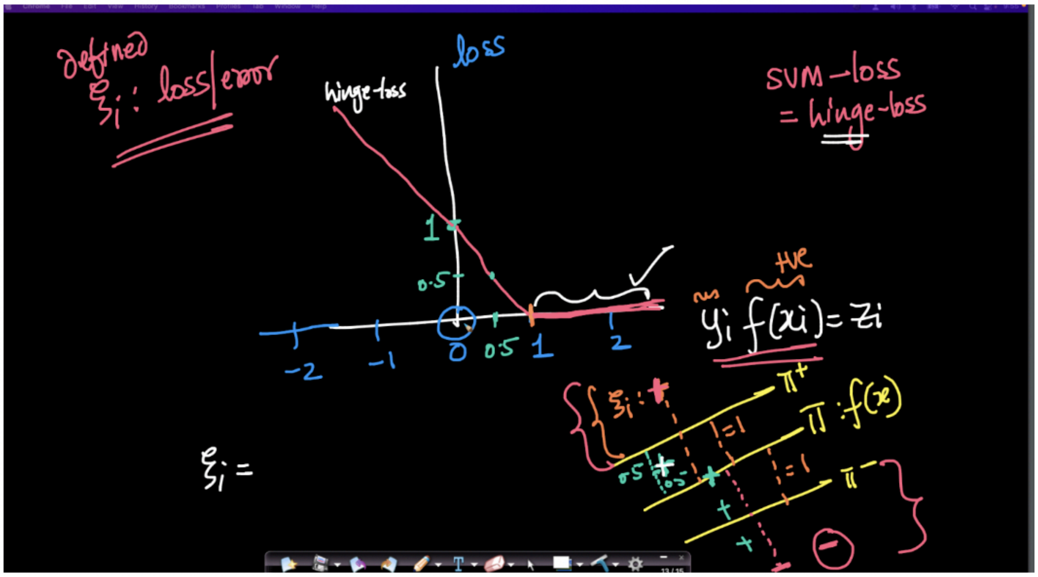
Where

- C is a hyperparameter that is analogous to the Regularization parameter (λ) and -
- $\frac{||w||}{2}$ is analogous to L2 Regularization.

Note:-

- This is the **Primal form** of SVM
- The primal form works similarly to Log Reg.

Understanding Hinge Loss



Dual form of loss in SVM

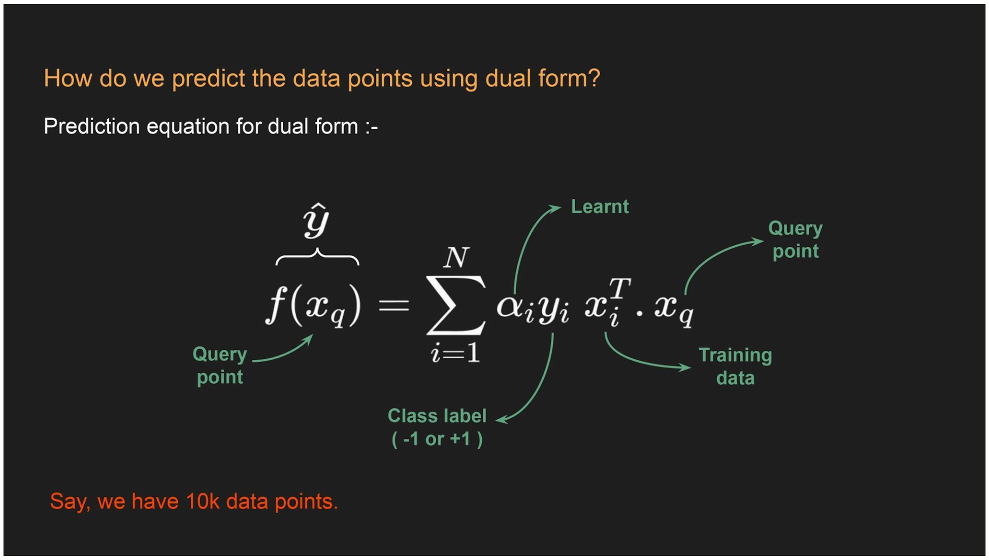
We define a variable $\propto_i$ for each data point x_i, such that

- $0 <= \propto_i <= C$
- $\sum_{i=1}^{n} \propto_i y_i = 0$

Dual form of loss function: $max_{\propto_i} \sum_{i=1}^{n} \propto_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \propto_i \propto_j y_i y_j x_i^T x_j$

Note: Dual form aids in the kernel trick by implicitly transforming the data points to a higher dimension

Prediction using dual form:



How do we predict the data points using dual form?

Prediction equation for dual form :-

$$\overbrace{f(x_q)}^{\hat{y}} = \sum_{i=1}^{N} \alpha_i y_i \; x_i^T . x_q$$

Learnt

Query point

Query point

Training data

Class label (-1 or +1)

Say, we have 10k data points.

How does the value of C affect the model?

C is the hyperparameter. It causes a tradeoff between maximizing the margin and minimizing ζ

- If C is very large,
    - The SVM model tries to minimize HingeLoss
    - This makes the model have 0 incorrect predictions.
    - Thus making the model **overfit**.
- If C is very small,
    - The model tends to generalize the data
    - Hence, the model tends to have a maximum ζ,
    - Thus making the model **underfit.**

What is the impact of Imbalanced Data on SVM?

SVM is impacted if there is an imbalance in Support Vectors. To resolve this:

- Either class weights should be used.
- Or rebalance the data.

What are some Limitations of the SVM model?

- In some situations, the RBF kernel is very similar to KNN.
- In practice, GBDT/Random Forest still beats SVM
- Time complexity to train SVM is very high: **O(n^2)**; n: no of data points

- Unlike Deep learning, SVM cannot create new features on its own.
- If we observe, we are just replacing feature engineering in GBDT/Random forest with kernel design in SVM.
- Even the RBF kernel SVM is impacted by outliers.

Support Vector Regressor (SVR)

Though not very popularly used, SVM can be used for regression problems also.

Loss function: $min_{(w,b)} \frac{1}{2} ||w||^2 + C. \in$**,** such that:

- $y_i - \hat{y}_i <= \in$
- $\hat{y}_i - y_i <= \in$
- $\in >= 0$