Lecture - 11 FP67Qd

FP67Qd

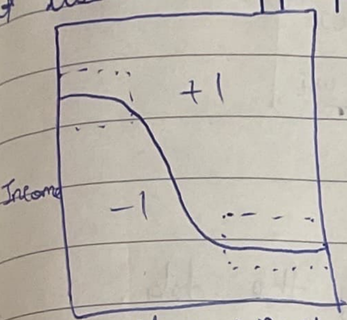
Recap: Linear model for 3 Learning problem.

$\boxed{h(x) = w^T x}$ → linear eqⁿ

* Linear Model has limitations:-

Linearly not seperable

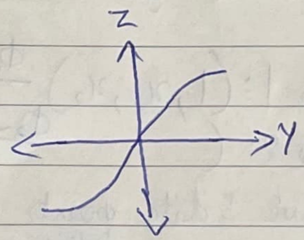
* Use the appropriate feature.



$Y >> 3$ years, no additional effect beyond $Y = 3$
$Y << 0.3$ years, no additional effect below $Y = 0.3$.

* Change the feature using a transform.

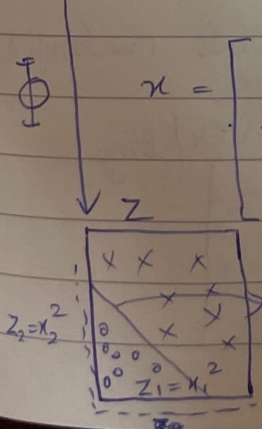
$Y \xrightarrow{\Phi} Z$



* Mechanics of Non-linear Feature transforms.

$(x - c_1)^2 + (y - c_2)^2 = r^2$, Since this is not linear. Lets perform feature transform.

$$x = \begin{bmatrix} 1 \\ x_1 \\ x_2 \end{bmatrix} \xrightarrow[\text{transform}]{\Phi} Z = \begin{bmatrix} 1 \\ z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} 1 \\ \Phi(x_1) \\ \Phi(x_2) \end{bmatrix} = \begin{bmatrix} 1 \\ x_1^2 \\ x_2^2 \end{bmatrix}$$



sign$(w^T Z)$

Feature transforms in ~~Genral~~ General.

x- space is R^d. Z space is $R^{\tilde{d}}$

$$x = \begin{bmatrix} 1 \\ x_1 \\ \vdots \\ x_d \end{bmatrix} \qquad z = \Phi(x) = \begin{bmatrix} 1 \\ \Phi_1(x) \\ \vdots \\ \Phi_{\tilde{d}}(x) \end{bmatrix} = \begin{bmatrix} 1 \\ z_1 \\ \vdots \\ z_{\tilde{d}} \end{bmatrix}$$

$(x_1, y_1), \ldots (x_N, y_N)$ $\qquad (z_1, y_1), \ldots (z_N, y_N)$

$g(x) = \text{sign}(\tilde{w}^T \Phi(x))$ $\qquad \tilde{w} = \begin{bmatrix} w_0 \\ w_1 \\ w_2 \\ \vdots \\ w_{\tilde{d}} \end{bmatrix}$

↑ Non linear eqn.

$\tilde{g}(z) = \text{sign}(\tilde{w}^T z)$.

↑ linear eqn.

* Do not pick Φ (feature transform after seeing the data. pick Φ before seeing the data.

* <u>Polynomial feature transform:-</u>

Degree 1: $(1, x_1, x_2) \xrightarrow{\Phi_1} (1, x_1, x_2)$. $\qquad \tilde{d}_{vc} = 3$

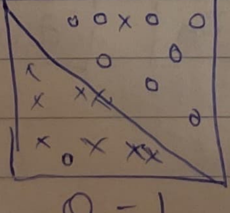
degree 2: $(\xrightarrow{\Phi_2} (1, x_1, x_2, x_1^2, x_1 x_2, x_2^2)$ $\qquad \tilde{d}_{vc} = 6$.

If we have 5 data points use 4th degree polynomial function.

$$x = \begin{bmatrix} 1 \\ x_1 \\ x_2 \end{bmatrix} \xrightarrow[\text{degree 2.}]{\Phi_2} \Phi_2(x) = \begin{bmatrix} 1 \\ \Phi_1(x) \\ \Phi_2(x) \\ \vdots \\ \Phi_5(x) \end{bmatrix} = \begin{bmatrix} 1 \\ x \\ x_2 \\ x^2 \\ x_1 x_2 \\ x_2^2 \end{bmatrix}$$

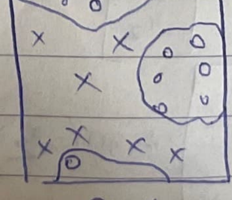
If we have higher order polynomial we will have less misclassified data points.

Take away:- Use simpler order function when we have small dataset



$\rho = 1$

overfitting.



$\rho = 4$.