

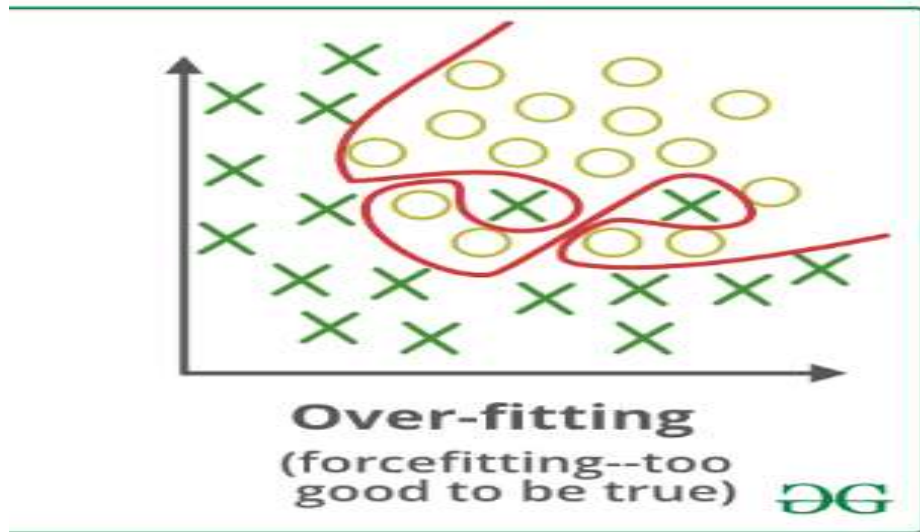
COURSE NAME: ARTIFICIAL INTELLIGENCE
COURSE CODE: CIS 412

SYED TANGIM PASHA
LECTURER,

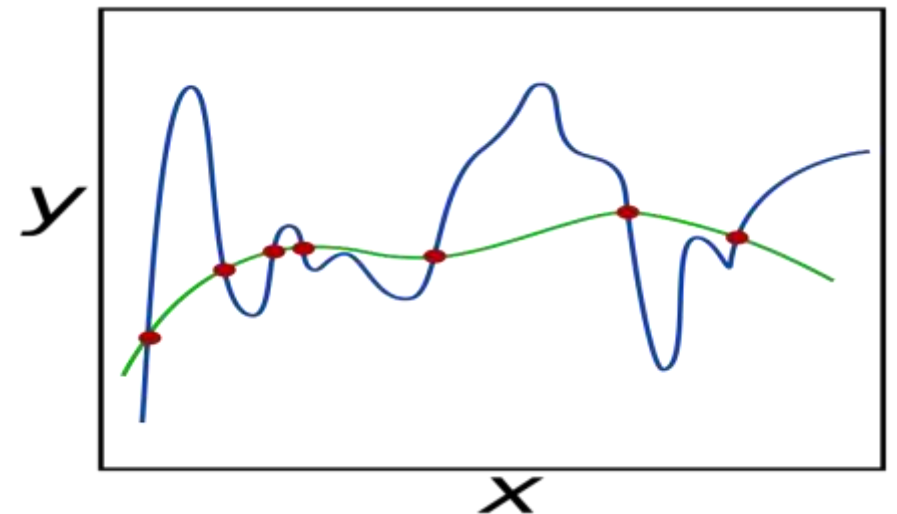
DEPARTMENT OF COMPUTING AND INFORMATION SYSTEM (CIS)
DAFFODIL INTERNATIONAL UNIVERSITY (DIU)
DHAKA, BANGLADESH

REGULARIZATION

- **Overfitting:** Overfitting is a phenomenon that occurs when a Machine Learning model is constraint to training set and not able to perform well on unseen data.
- **Regularization:** Regularization is a technique used to reduce the errors by fitting the function appropriately on the given training set and avoid overfitting.



OVERFITTING PROBLEM



REGULARIZATION PARAMETER-LAMBDA (λ)

- **LAMBDA:** The lambda parameter controls the amount of regularization applied to the model.
- If λ is too large, it causes Underfitting.
- If λ is 0 or too small, it causes Overfitting.
- λ is a **hyperparameter**.

REGULARIZATION (TYPES)

- L1 Regularization
- L2 Regularization
- Dropout Regularization

L1 REGULARIZATION

- **L1 Regularization:** L1 regularization, also known as **L1 norm** or **Lasso regression**, combats overfitting by shrinking the parameters towards 0.

$$\frac{\lambda}{2m} \sum_{j=1}^n |w_j|$$

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x^i) - y^i)^2 + \frac{\lambda}{2m} \sum_{j=1}^n \theta_j^2$$

LINEAR REGRESSION COST FUNCTION AFTER ADDING L1 REGULARIZATION

L2 REGULARIZATION

- **L2 Regularization:** l2 regularization, or **L2 norm**, or **Ridge regression** combats overfitting by forcing weights to be small, but not making the exactly 0.

$$\frac{\lambda}{2m} \sum_{j=1}^n w_j^2$$

$$\frac{1}{2m} \sum_{i=1}^m \{h_w(x^{(i)}) - y^{(i)}\}^2 + \frac{\lambda}{2m} \sum_{j=1}^n w_j^2$$

LINEAR REGRESSION COST FUNCTION AFTER ADDING L2 REGULARIZATION

SPARSE MATRIX & SPARSITY

- **Sparsity:** In AI and Machine Learning, sparsity refers to a matrix of numbers that includes many zeros or values that will not significantly impact a calculation.
- **Sparse Matrix:** Sparse matrices are those matrices that have the majority of their elements equal to zero. In other words, the sparse matrix can be defined as the matrix that has a greater number of zero elements than the non-zero elements. **Sparse matrix benefits:**

(1) **Storage:** We know that a sparse matrix contains lesser non-zero elements than zero. So less memory can be used to store elements. It evaluates only the non-zero elements.

(2) **Computing Time:** In the case of searching sparse matrix, we need to traverse only the non-zero elements rather than traversing all the sparse matrix elements. It saves computing time by logically designing a data structure traversing non-zero elements.

Sparse Matrix →

	0	1	2	3
0	0	4	0	5
1	0	0	3	6
2	0	0	2	0
3	2	0	0	0
4	1	0	0	0

L1 vs L2 REGULARIZATION

L1 REGULARIZATION	L2 REGULARIZATION
(1) Penalizes the sum of absolute value of weights.	(1) Penalizes the sum of square weights.
(2) It has a sparse solution.	(2) It has a non-sparse solution.
(3) Constructed in feature selection.	(3) No feature selection.
(4) Robust to outliers.	(4) Not robust to outliers.