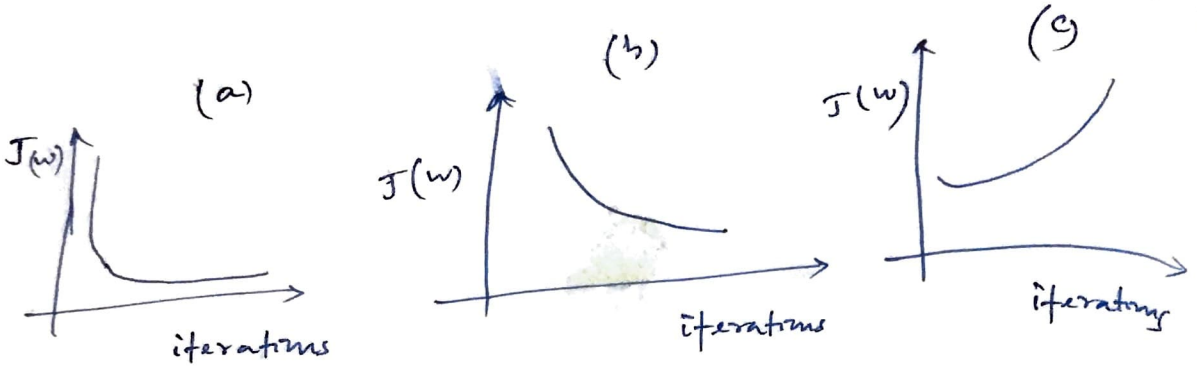


* In case of high learning rate, step will be very high. ~~As~~

$$w_j^{t+1} = w_j^t - \alpha \sum_{i=1}^{m} \left(h_w(x^{(i)}) - y^{(i)} \right) x_j^{(i)}$$

$$x^{(i)} = \begin{bmatrix} x_0^{(i)} & x_1^{(i)} & x_2^{(i)} & \cdots & x_n^{(i)} \end{bmatrix}$$

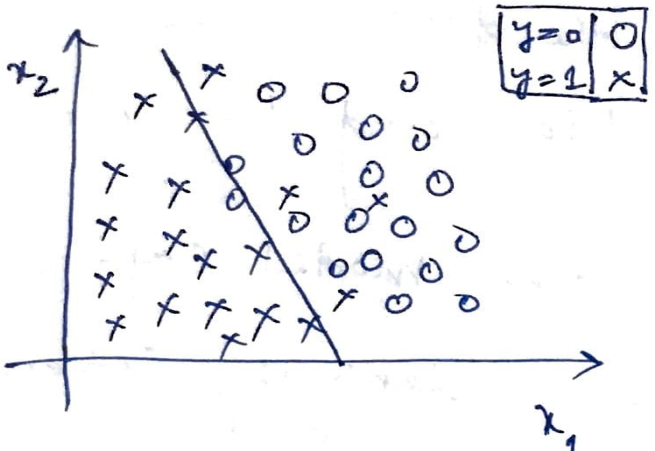
* The objective function will decrease quickly initially, but it will not find the global minima and objective function starts increasing after a few iterations.

* In case of low learning rate, the step will be small. So, the objective or cost function will decrease slowly.

$$h_w(x) = w_0 + w_1 x_1 + w_2 x_2$$

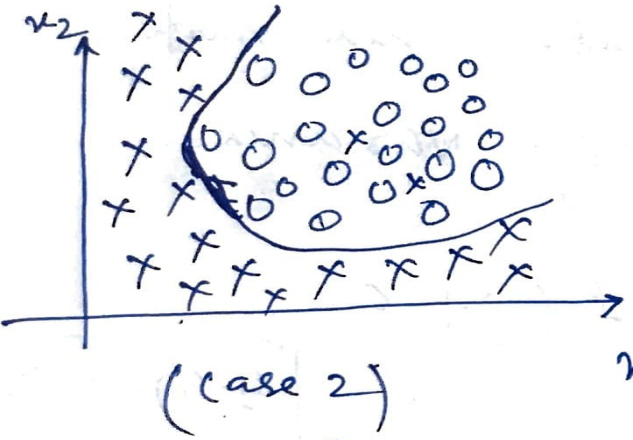
Polynomial regression: $h_w(x) = w_0 + w_1 x_1 + w_2 x_1^2 + w_2 x_1^3 + w_2 x_2 + w_2 x_2^2 + w_3 x_2^3$

Bias Variance tradeoff



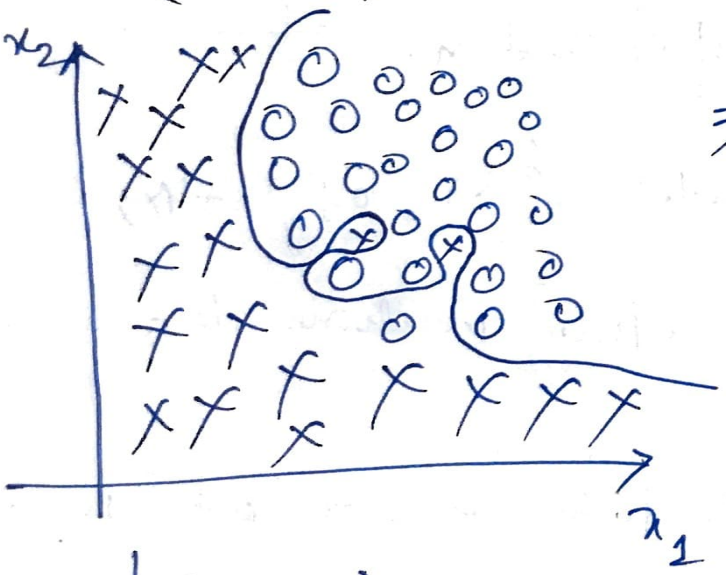
(Case 1)

⇒ In case 3, the decision boundary is not smooth this means it will over-fitting the data.



(case 2)

⇒ In case 3, a higher degree & polynomial might have a very high accuracy on the training data but is expected to fail badly on test dataset



(case 3)

⇒ In case 1, the training misclassification rate is high because it underfits the training data.

⇒ In case 2, there is a tradeoff between bias and variance.

⇒ underfitting happens when the model unable to capture the underlying pattern of the data. These models usually have high bias and low variance

⇒ overfitting happens when the model captures the noise along with the underlying pattern in data. It happens when we train our model a lot over noisy dataset. These models have low bias and high variance

$$Y = f(X) + \epsilon$$

$$f(x) = h_w(x)$$

the ϵ is Gaussian distribution.

we may estimate a model $\hat{f}(x)$ of $f(x)$ using linear regression or any other machine learning method.

The expected square prediction error at point x is

$$err(x) = E\left[Y - \hat{f}(x)\right]$$

The $err(x)$ has two components as bias and variance.

$$err(x) = \left(E\left[\hat{f}(x)\right] - f(x)\right)^2 + E\left[\hat{f}(x) - E\left[\hat{f}(x)\right]\right]^2 + \sigma_e^2$$

$$err(x) = Bias^2 + Variance + Irreducible\ error$$

The third term, irreducible error is the noise term in the true relationship that cannot fundamentally be reduced by any model.

error due to bias

* The error due to bias is taken as the difference between the expected (or average) prediction of our model and the correct value which we are trying to predict.

* Bias measures how far off in general these models predictions are from the correct value.

Error due to variance

The error due to variance is taken as the variability of a model prediction for a given data point. Again, imagine you can repeat the entire model building process multiple times. The variance is how much the predictions for a given point vary between different realizations of the model.