Proof of Gradient Descent Convergence :-

Let there be a cost function $C(\vec{X})$ where $\vec{X} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$

Therefore, by definition of Gradient Descent

$$X_1 = X_0 - \alpha \nabla C_0 \quad -(1)$$ (The subscript represents the iteration number)

where α is the learning rate (hyperparameter) and $\nabla C = \begin{bmatrix} \frac{\partial C}{\partial x_1} \\ \frac{\partial C}{\partial x_2} \\ \vdots \\ \frac{\partial C}{\partial x_n} \end{bmatrix}$

Using (1),

$$C(\vec{X_1}) = C(\vec{X_0} - \alpha \nabla C_0)$$

$$= C(\vec{X_0}) - \alpha \nabla C_0 (\nabla C_0)$$ [Taylor's expansion in linear order]

$$\therefore C(\vec{X_1}) = C(\vec{X_0}) - \alpha |\nabla C_0|^2$$

$|||^{ly}$ $$C(\vec{X_2}) = C(\vec{X_1}) - \alpha |\nabla C_1|^2$$

$$\vdots$$

$$C(\vec{X_n}) = C(X_{n-1}) - \alpha |\nabla C_{n-1}|^2$$

$\therefore$ Adding all the eqns,

$$C(\vec{X_n}) = C(\vec{X_0}) - \alpha \left(\sum_{i=0}^{n-1} |\nabla C_i|^2\right)$$

As n tends to ∞, we can see that the summation in the R.H.S. is an infinite sum, for infinite sums to converge to a value, $\lim_{k\to\infty} \nabla C_k = 0$ (Contrapositive of divergence test statement).

Therefore, if our cost function is bounded from below, $C(\vec{X_0}) + C(\vec{X_n})$ is finite. Therefore, $\sum_{i=0}^{n-1} |\nabla C_i|^2$ is finite. Therefore, as $\lim_{k\to\infty} \nabla C_k = 0$. This shows that our gradient becomes zero at sufficiently high number of iterations.

Also, $\because \lim_{k\to\infty} \nabla C_k = 0$,

$$\lim_{k\to\infty} C(\vec{X_{k+1}}) = C(\vec{X_k}) - \alpha |\nabla C_k|^2 \approx C(\vec{X_k})$$

Therefore, cost function reaches a minimum which is stable. (as $C(\vec{x_n}) \leq C(\vec{x_{n-1}}) \leq \cdots \leq C(\vec{x_1}) \leq C(\vec{x_0})$)