Perceptron Convergence theorem: If samples given are linearly separable, then the fixed-increment perceptron algorithm terminates after a finite number of weight vector updates.

Proof: we can prove this theorem by showing the number updations on weight vector is bounded by a finite number.
For ease of computation let us make some assumptions without loosing any generality.
Assumption 1: Take zero vector (all elements are zero valued) as initial weight vector instead of choosing it randomly.
Assumption 2: Learning rate is fixed and valued as unity.
Assumption 3: All feature vectors are augmented.
Assumption 4: As we have two classes {+1,-1}, multiply the feature vectors by -1, whose class is -1. so that a weight vector W represents a separating hyperplane if $W^T.X_i > 0$, $\forall$ i

So the new update rule can be written as

$$W(k+1) = W(k) + X(K) \text{ , } \forall \text{ k such that } W(k)^T.X(k) \leq 0$$

from above

$$\begin{aligned} \| W(K+1) \|^2 &= \| W(k) + X(k) \|^2 \\ &= \| W(k) \|^2 + \| X(k) \|^2 + 2\, W(k)T\, X(k) \\ &\leq \| W(k) \|^2 + \| X(k) \|^2 \qquad \text{(because } W(k)^T.X(k) \leq 0 \text{ for all misclassified samples)} \end{aligned}$$

If we recurse on it we can get

$$\| W(K+1) \|^2 \leq \|W(0)\|^2 + \sum_{i=o}^{k} \| X(i) \| 2$$

Here, W(0) is zero vector from our assumption 1 and let M is MAX_i ($\|$ X(i) $\|^2$) , then above equation becomes $\| W(K+1) \|^2 \leq kM$ -----(1)

From update rule we can say that W(K+1) is linear combination of misclassified samples feature vectors and initial weight vector i.e $W(k+1) = W(0) + \sum_{i=0}^{k} X(i)$

$$W(k+1) = \sum_{i=0}^{k} X(i) \quad \text{------(2)}$$

Now consider a weight vector W* from solution space such that $X_i^T W^* > 0$,$\forall$ i (because W* separates all samples correctly)

consider $N = MIN_i\ (X_i^T W^*)$, note that N>0

from equation 2 we have

$$W(K+1).W^* = \sum_{i=0}^{k} X(i) \geq kN > 0$$

$$k^2 N^2 \leq \| W(K+1).W^* \|^2$$

From Cauchy - Schwarz inequality

$$K^2N^2 \leq \| W(K+1) \|^2 \, \| W^* \|^2$$
$$\leq kM \,.\, \| W^* \|^2 \qquad \text{from(1)}$$

i.e $k \leq \| W^*\|^2 \,.\, M / N^2$

Here W^* is finite vector if given samples are linearly separable and from above M and N are finite values obtained from feature vectors(samples)

So k is bounded by a finite value, from this we can infer that perceptron algorithm stops after some finite number of steps i.e Perceptron algorithm finds a separating hyperplane with in the finite number of steps.