Recap (Theory)

Let us quickly recap various components in the general settings:

1. **Training data**: (features, label) or $(\mathbf{X}, y)$, where label y is a **discrete** number from a finite set.**Features** in this case are pixel values of an image.
2. **Model :**

$$\begin{aligned} h_{\mathbf{w}} : y &= \mathrm{g}(\mathbf{w}^T \mathbf{x}) \\ &= \mathrm{g}(w_0 + w_1 x_1 + \ldots + w_m x_m) \end{aligned}$$

   where,

   - $\mathbf{w}$ is a weight vector in $\mathbb{R}^{(m+1)}$ i.e. it has components: $\{w_0, w_1, \ldots, w_m\}$
   - $\mathrm{g}(.)$ is a non-linear activation function given by a signum function:

$$\mathrm{g}(z) = \begin{cases} +1, & \text{if } z \geq 0 \\ -1, & \text{otherwise (i.e. } z < 0) \end{cases}$$

3. **Loss function:** Let $\widehat{y^{(i)}} \in \{-1, +1\}$ be the prediction from perceptron and $y^{(i)}$ be the actual label for i-th example. The error is

$$e^{(i)} = \begin{cases} 0, & \text{if } \widehat{y^{(i)}} = y^{(i)} \\ -\mathbf{w}^T \mathbf{x}^{(i)} y^{(i)}, & \text{otherwise (i.e. } \widehat{y^{(i)}} \neq y^{(i)}) \end{cases}$$

   This can be compactly written as:

$$e^{(i)} = \max(-\mathbf{w}^T \mathbf{x}^{(i)} y^{(i)}, 0) = \max(-h_{\mathbf{w}}(\mathbf{x}^{(i)}) y^{(i)}, 0)$$

4. **Optimization:**

   - Perceptron learning algorithm

   1. Initialize $\mathbf{w}^{(0)} = \mathbf{0}$
   2. For each training example $\left(\mathbf{x}^{(i)}, y^{(i)}\right)$:
      - $\hat{y}^{(i)} = \text{sign}\left(\mathbf{w}^T \mathbf{x}^{(i)}\right)$ [Calculate the output value]
      - $\mathbf{w}^{(t+1)} := \mathbf{w}^{(t)} + \alpha(y^{(i)} - \hat{y}^{(i)})\, \mathbf{x}^{(i)}$ [Update the weights]

> Linearly separable examples lead to convergence of the algorithm with zero training loss, else it oscillates.

Parameters of Perceptron class

- Let's quickly take a look into the important parameters of the Perceptron()

```
class sklearn.linear_model.Perceptron(*, penalty=None, alpha=0.0001, l1_ratio=0.15, fit_intercept=True, max_iter=1000, tol=0.001, shuffle=True, verbose=0, eta0=1.0, n_jobs=None, random_state=0, early_stopping=False, validation_fraction=0.1, n_iter_no_change=5, class_weight=None, warm_start=False).
```