Weight shrinkage.

We do Ridge regression:

$$\nabla \mathcal{L} = \frac{1}{N} \sum_{n}^{N} (\vec{w} \vec{x}_n - t_n) x_n + \lambda \vec{w}$$

So the basic GD step is:

$$\theta \mapsto \theta - \eta \nabla \mathcal{L}$$

or $$\vec{w} \mapsto \vec{w} - \eta \lambda \vec{w} + -\eta \frac{1}{N} \sum_{n=1}^{N} (w x_n - t_n) x_n$$

$$\vec{w}(\underbrace{1 - \eta \lambda}) - \underbrace{\qquad\qquad}$$

geometrical decay with a rate $1 - \eta\lambda < 1$.

normal MSE term.

Hence the name "weight decay"

or "parameter shrinkage"

(a geometric series decay exponentially fast, so if the $-\frac{\eta}{N} \sum_{n}^{N} ()_n$ term does not push $\vec{w}$, it goes towards 0).