algorithm terminates when no parameters have improved over the best recorded validation error for some pre-specified number of iterations. This procedure is specified more formally in algorithm 7.1.

Algorithm 7.1 The early stopping meta-algorithm for determining the best amount of time to train. This meta-algorithm is a general strategy that works well with a variety of training algorithms and ways of quantifying error on the validation set.

Let n be the number of steps between evaluations.

Let p be the "patience," the number of times to observe worsening validation set error before giving up.

Let $\boldsymbol{\theta}_o$ be the initial parameters.

$\boldsymbol{\theta} \leftarrow \boldsymbol{\theta}_o$

$i \leftarrow 0$

$j \leftarrow 0$

$v \leftarrow \infty$

$\boldsymbol{\theta}^* \leftarrow \boldsymbol{\theta}$

$i^* \leftarrow i$

while $j < p$ **do**

    Update $\boldsymbol{\theta}$ by running the training algorithm for n steps.

    $i \leftarrow i + n$

    $v' \leftarrow \text{ValidationSetError}(\boldsymbol{\theta})$

    if $v' < v$ **then**

        $j \leftarrow 0$

        $\boldsymbol{\theta}^* \leftarrow \boldsymbol{\theta}$

        $i^* \leftarrow i$

        $v \leftarrow v'$

    else

        $j \leftarrow j + 1$

    end if

end while

Best parameters are $\boldsymbol{\theta}^*$, best number of training steps is i^*

This strategy is known as **early stopping**. It is probably the most commonly used form of regularization in deep learning. Its popularity is due both to its effectiveness and its simplicity.

One way to think of early stopping is as a very efficient hyperparameter selection algorithm. In this view, the number of training steps is just another hyperparameter. We can see in figure 7.3 that this hyperparameter has a U-shaped validation set