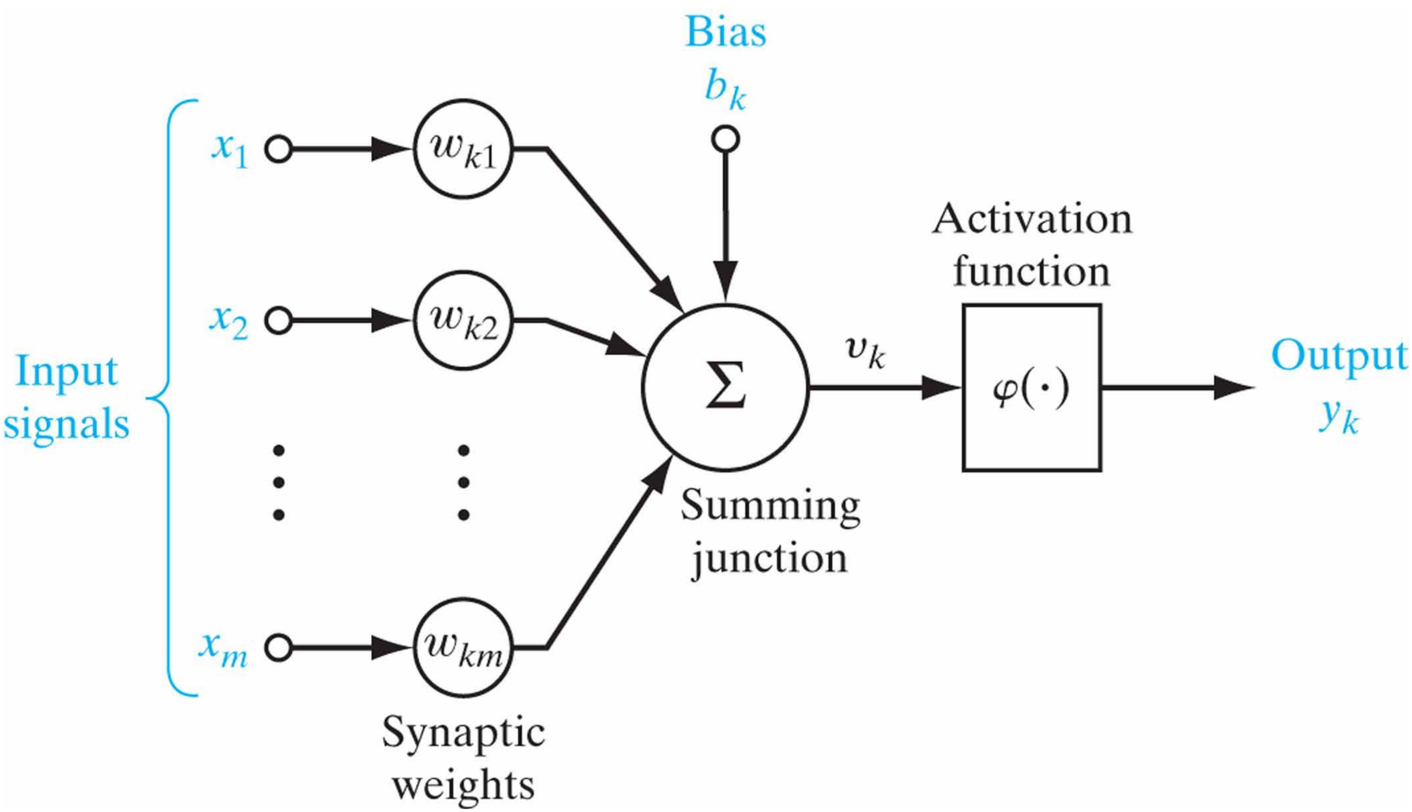
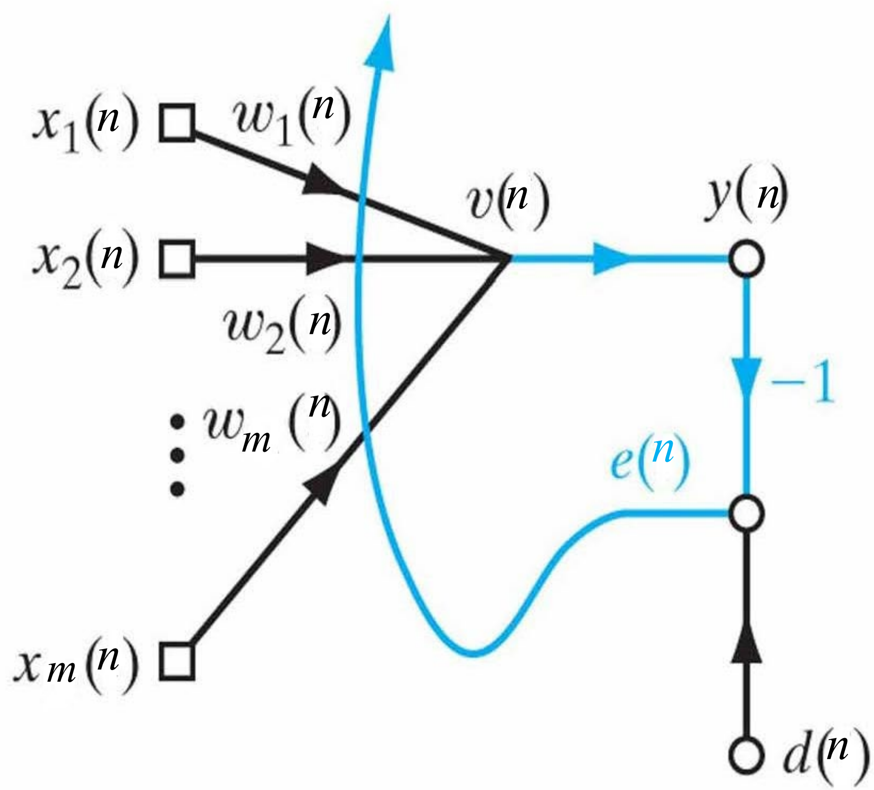
Model of a single neuron:

Perceptron (a nonlinear neuron)

LMS (a linear neuron)





$x_1(n)$
$w_1(n)$
$v(n)$
$y(n)$
$x_2(n)$
$w_2(n)$
-1
$w_m(n)$
$e(n)$
$x_m(n)$
$d(n)$

The LMS algorithm is configured to minimize the instantaneous value of the cost function ,

$$\mathcal{E}(\hat{\mathbf{w}}) = \frac{1}{2} e^2(n)$$

where $e(n)$ is the error signal measured at time n,

$$e(n) = d(n) - \mathbf{x}^T(n)\hat{\mathbf{w}}(n)$$

Differentiating $\mathcal{E}(\hat{\mathbf{w}})$ with respect to the weight vector $\hat{\mathbf{w}}$ yields

$$\frac{\partial \mathcal{E}(\hat{\mathbf{w}})}{\partial \hat{\mathbf{w}}} = e(n) \frac{\partial e(n)}{\partial \hat{\mathbf{w}}(n)}$$

where

$$\frac{\partial e(n)}{\partial \hat{\mathbf{w}}(n)} = -\mathbf{x}(n)$$

Thus

$$\hat{\mathbf{g}}(n) = \frac{\partial \mathcal{E}(\hat{\mathbf{w}})}{\partial \hat{\mathbf{w}}(n)} = -\mathbf{x}(n)e(n)$$

The LMS algorithm

$$\boxed{\hat{\mathbf{w}}(n+1) = \hat{\mathbf{w}}(n) + \eta\mathbf{x}(n)e(n)}$$

The LMS Algorithm

Training Sample: Input signal vector $= \mathbf{x}(n)$

Desired response $= d(n)$

User-selected parameter: η

Initialization. Set $\hat{\mathbf{w}}(0) = \mathbf{0}$.

Computation. For $n = 1, 2, \ldots,$ compute

$$e(n) = d(n) - \hat{\mathbf{w}}^T(n)\mathbf{x}(n)$$

$$\hat{\mathbf{w}}(n + 1) = \hat{\mathbf{w}}(n) + \eta\mathbf{x}(n)e(n)$$

Remarks on the LMS

- LMS is a stochastic gradient algorithm
- As iteration number n increases, $\hat{w}(n)$ performs a random walk (Brownian motion) about the Wiener solution w_0 - LMS does not require statistics from the environment
- Langevin force responsible for the none-equilibrium behavior of LMS
- One key assumption in LMS convergence is small learning rate η (condition needed for analysis using Kushner's direct-averaging method)
- LMS convergence properties sensitive to the condition number of R_{xx}
- Generally LMS converges slowly

The Least-Mean Square (LMS) Algorithm

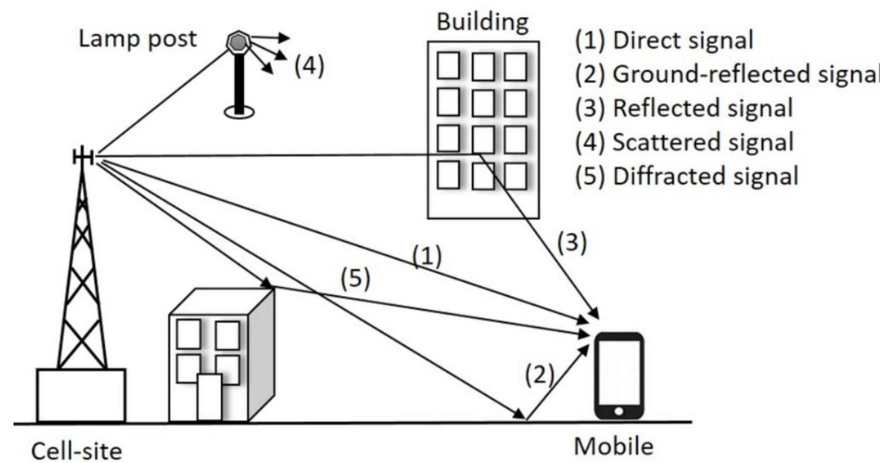
– Widrow & Hoff 1960

- Inspired by perceptron, linear combiner
- The first linear adaptive filtering algorithm
- Adaptive version of “Wiener filter” (some properties)
- “linear” complexity, yet very effective in applications, simple to code, robust to disturbances
- It is a stochastic gradient algorithm
- Set the stage for backpropagation

The Least-Mean Square (LMS) Algorithm

– Widrow & Hoff 1960

- Echo cancellation, noise cancellation, channel equalization in communication, system ID…



Multipath propagation outdoor scenario (modeled as time varying finite duration impulse response filter

Echo cancellation