Logistic regression - Outline

- The model
- Error measure
- Learning algorithm
How to minimize $E_{\text{in}}$

For logistic regression,

$$E_{\text{in}}(w) = \frac{1}{N} \sum_{n=1}^{N} \ln \left( 1 + e^{-y_n w^T x_n} \right) \quad \leftarrow \text{iterative solution}$$

Compare to linear regression:

$$E_{\text{in}}(w) = \frac{1}{N} \sum_{n=1}^{N} (w^T x_n - y_n)^2 \quad \leftarrow \text{closed-form solution}$$
Iterative method: gradient descent

General method for nonlinear optimization

Start at $\mathbf{w}(0)$; take a step along steepest slope

Fixed step size: $\mathbf{w}(1) = \mathbf{w}(0) + \eta \hat{\mathbf{v}}$

What is the direction $\hat{\mathbf{v}}$?
Formula for the direction $\hat{v}$

$$
\Delta E_{in} = E_{in}(w(0) + \eta \hat{v}) - E_{in}(w(0))
$$

$$
= \eta \nabla E_{in}(w(0))^T \hat{v} + O(\eta^2)
$$

$$
\geq -\eta \| \nabla E_{in}(w(0)) \|
$$

Since $\hat{v}$ is a unit vector,

$$
\hat{v} = - \frac{\nabla E_{in}(w(0))}{\| \nabla E_{in}(w(0)) \|}
$$
Fixed-size step?

How $\eta$ affects the algorithm:

$\eta$ too small

$\eta$ too large

variable $\eta$ – just right

$\eta$ should increase with the slope
Easy implementation

Instead of

$$\Delta w = \eta \hat{v}$$

$$= - \eta \frac{\nabla E_{in}(w(0))}{||\nabla E_{in}(w(0))||}$$

Have

$$\Delta w = - \eta \nabla E_{in}(w(0))$$

Fixed learning rate $\eta$
Logistic regression algorithm

1. Initialize the weights at $t = 0$ to $\mathbf{w}(0)$
2. for $t = 0, 1, 2, \ldots$ do
3. Compute the gradient

$$\nabla E_{\text{in}} = -\frac{1}{N} \sum_{n=1}^{N} \frac{y_n x_n}{1 + e^{y_n \mathbf{w}^T(t) x_n}}$$

4. Update the weights: $\mathbf{w}(t + 1) = \mathbf{w}(t) - \eta \nabla E_{\text{in}}$
5. Iterate to the next step until it is time to stop
6. Return the final weights $\mathbf{w}$