1 Logistic Regression

1.1 Batch Gradient Descent

a.

$$NLL\,(w) = -\sum_{i=1}^{N}\Big[(1-y_i)\log(1-\sigma(\mathbf{w}^T\mathbf{x})) + y_i\log\sigma(\mathbf{w}^T\mathbf{x})\Big]$$

$$= -\sum_{i=1}^{N}[y_i\mathbf{w}^T\mathbf{x} - \log(1+e^{\mathbf{w}^T\mathbf{x}})]$$

$$\frac{\delta NLL(w)}{\delta \mathbf{w}} = -\sum_{i=1}^{N}\mathbf{x}(y_i - \sigma(\mathbf{w}^T\mathbf{x}))$$

1.2 Stochastic Gradient Descent

a.

$$l(w) = \Big[(1-y_t)\log(1-\sigma(\mathbf{w}^T\mathbf{x_t})) + y_t\log\sigma(\mathbf{w}^T\mathbf{x_t})\Big]$$

b.

$$w_t = w_t - \eta((y_t - \sigma(\mathbf{w}^T\mathbf{x_t}))x_t)$$

c.

$$O(n)$$

d.

Larger values make the ROC curve closer and closer to the dotted line where as smaller values push it farther away

e.

$$w_t = w_t - \eta((y_t - \sigma(\mathbf{w}^T\mathbf{x_t}))x_t - \mu\|\mathbf{w}\|^2)$$

2

2.4 c.

The two curves appear to be the same