CIS 678 – Machine Learning

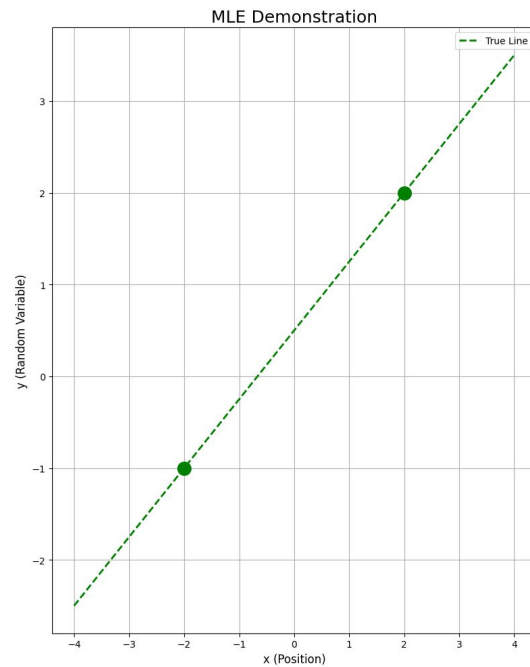- Maximum Likelihood Learning

Linear Regression: Probabilistic Twin

Method of Least Squares
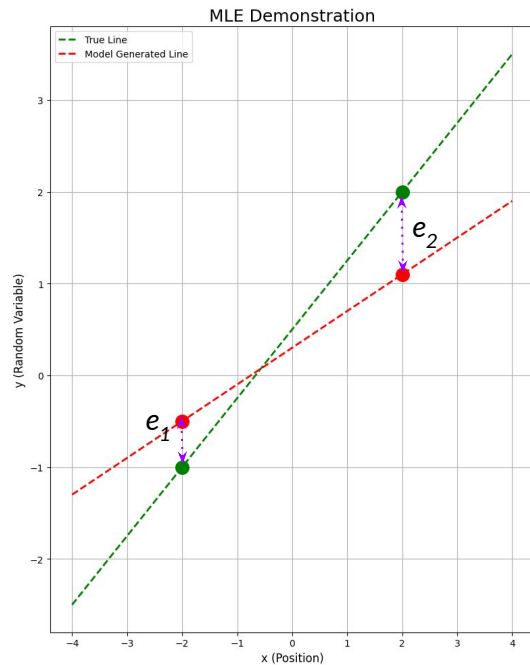
$$E(\mathbf{w}) = \frac{1}{2} \sum_{n=1}^{N} \{y(x_n, \mathbf{w}) - t_n\}^2$$

Linear Regression

MLE Demonstration

Least Squares Solution

$$E(\mathbf{w}) = \frac{1}{2} \sum_{n=1}^{N} \{y(x_n, \mathbf{w}) - t_n\}^2$$

MLE Demonstration

Normal (Gaussian) Distribution

▶ **Definition:** A continuous, symmetric, bell-shaped probability distribution.

▶ **Applications:** Test scores, heights, errors, finance, etc.

▶ **Parameters:**
  ▶ Mean (μ): center of the distribution
  ▶ Standard deviation (σ): spread of the data

▶ **Empirical Rule:**
  ▶ 68% within $\mu \pm 1\sigma$
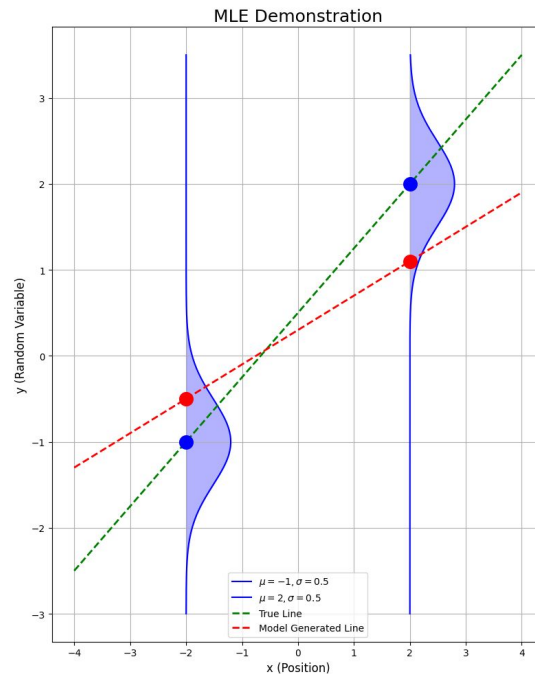 ▶ 95% within $\mu \pm 2\sigma$
  ▶ 99.7% within $\mu \pm 3\sigma$

"Empirical Rule"
approximations

68%

95%

99.7%

MathBits

μ-3σ μ-2σ μ-1σ μ μ+1σ μ+2σ μ+3σ

Probability density function	$f(x) = \dfrac{1}{\sigma\sqrt{2\pi}}\, e^{\frac{-(x-\mu)^2}{2\sigma^2}}$
parameters	$\mu =$ mean of x $\sigma =$ standard deviation of x

$\pi \approx 3.14159 \ldots$
$e \approx 2.71828 \ldots$

Probabilistic Twin

$$p(\mathbf{t}|\mathbf{x}, \mathbf{w}, \beta) = \prod_{n=1}^{N} \mathcal{N}\left(t_n | y(x_n, \mathbf{w}), \beta^{-1}\right).$$

MLE Demonstration

Probabilistic Twin

Probabilistic Formulation: Modeling Error Distribution

$$p(\mathbf{t}|\mathbf{x}, \mathbf{w}, \beta) = \prod_{n=1}^{N} \mathcal{N}\left(t_n | y(x_n, \mathbf{w}), \beta^{-1}\right).$$

MLE Demonstration

Probabilistic Twin

Probabilistic Formulation: Modeling Error Distribution

$$p(\mathbf{t}|\mathbf{x}, \mathbf{w}, \beta) = \prod_{n=1}^{N} \mathcal{N}\left(t_n | y(x_n, \mathbf{w}), \beta^{-1}\right).$$

$$= \prod_{n=1}^{N} \frac{1}{\sqrt{2\pi\beta^{-2}}} \exp\left\{-\frac{1}{2\beta^{-2}}\{y(x_n, \mathbf{w}) - t_n\}^2\right\}$$
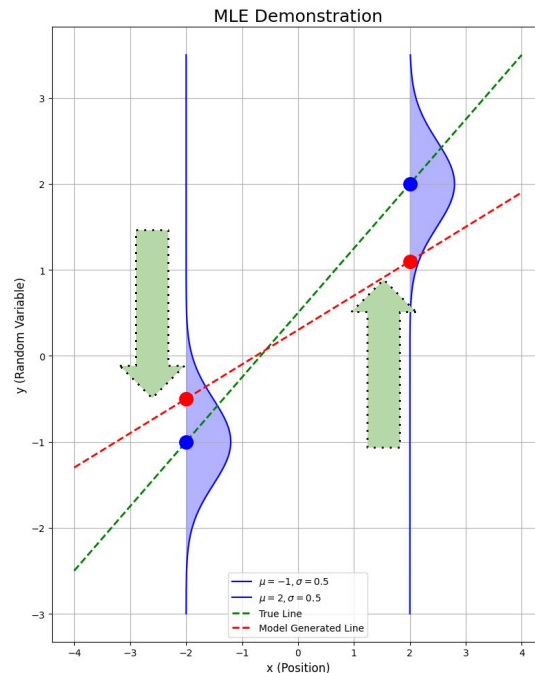

MLE Demonstration

Probabilistic Twin

Probabilistic Formulation: Modeling Error Distribution

$$p(\mathbf{t}|\mathbf{x}, \mathbf{w}, \beta) = \prod_{n=1}^{N} \mathcal{N}\left(t_n | y(x_n, \mathbf{w}), \beta^{-1}\right).$$

$$= \prod_{n=1}^{N} \frac{1}{\sqrt{2\pi\beta^{-2}}} \exp\left\{-\frac{1}{2\beta^{-2}}\{y(x_n, \mathbf{w}) - t_n\}^2\right\}$$

Taking the log

$$\ln p(\mathbf{t}|\mathbf{x}, \mathbf{w}, \beta) = -\frac{\beta}{2}\sum_{n=1}^{N}\{y(x_n, \mathbf{w}) - t_n\}^2 + \frac{N}{2}\ln\beta - \frac{N}{2}\ln(2\pi).$$

MLE Demonstration

Probabilistic Twin

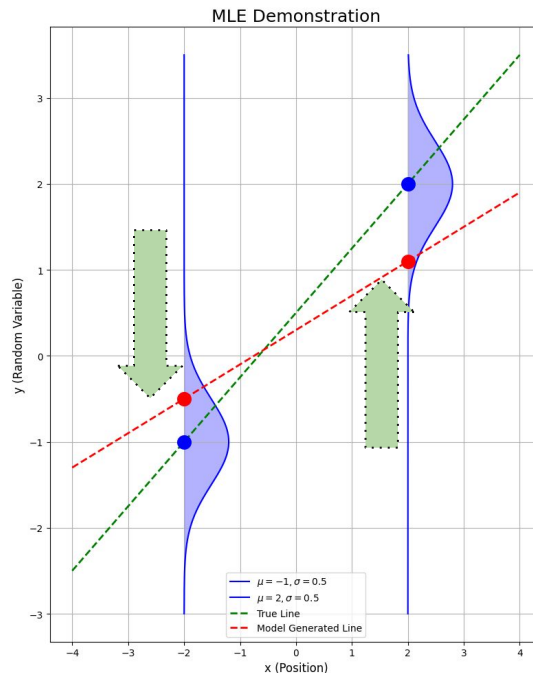**Probabilistic Formulation: Modeling Error Distribution**

$$p(\mathbf{t}|\mathbf{x}, \mathbf{w}, \beta) = \prod_{n=1}^{N} \mathcal{N}\left(t_n | y(x_n, \mathbf{w}), \beta^{-1}\right).$$

$$= \prod_{n=1}^{N} \frac{1}{\sqrt{2\pi\beta^{-2}}} \exp\left\{-\frac{1}{2\beta^{-2}}\{y(x_n, \mathbf{w}) - t_n\}^2\right\}$$

Taking the log

$$\ln p(\mathbf{t}|\mathbf{x}, \mathbf{w}, \beta) = -\frac{\beta}{2}\sum_{n=1}^{N}\{y(x_n, \mathbf{w}) - t_n\}^2 + \frac{N}{2}\ln\beta - \frac{N}{2}\ln(2\pi).$$

It's called Log likelihood!

MLE Demonstration

Probabilistic Twin
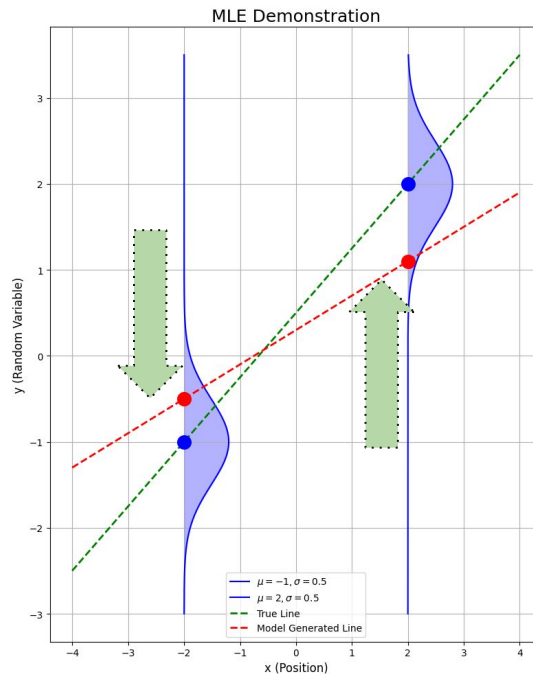
Probabilistic Formulation: Modeling Error Distribution

It's called Log likelihood!

$$\ln p(\mathbf{t}|\mathbf{x}, \mathbf{w}, \beta) = -\frac{\beta}{2} \sum_{n=1}^{N} \{y(x_n, \mathbf{w}) - t_n\}^2 + \frac{N}{2} \ln \beta - \frac{N}{2} \ln(2\pi).$$

Does it look familiar???

$$E(\mathbf{w}) = \frac{1}{2} \sum_{n=1}^{N} \{y(x_n, \mathbf{w}) - t_n\}^2$$

MLE Demonstration

Probabilistic Twin

Probabilistic Formulation: Modeling Error Distribution
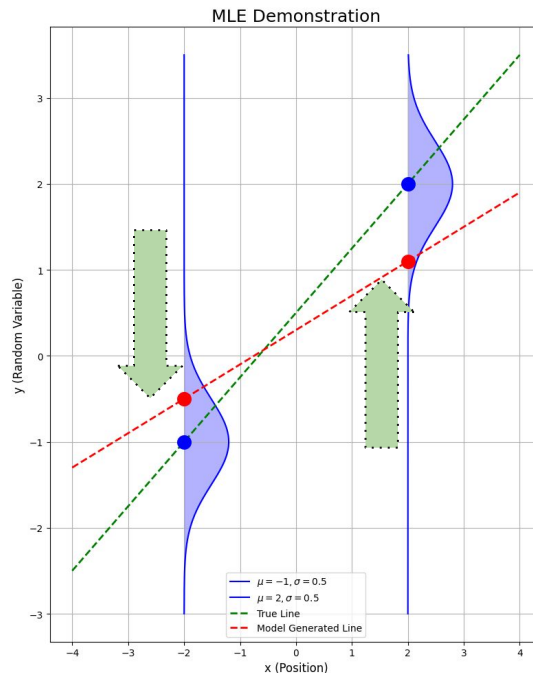
It's called Log likelihood!

$$\ln p(\mathbf{t}|\mathbf{x}, \mathbf{w}, \beta) = -\frac{\beta}{2} \sum_{n=1}^{N} \{y(x_n, \mathbf{w}) - t_n\}^2 + \frac{N}{2} \ln \beta - \frac{N}{2} \ln(2\pi).$$

Does it look familiar???

$$E(\mathbf{w}) = \frac{1}{2} \sum_{n=1}^{N} \{y(x_n, \mathbf{w}) - t_n\}^2$$

Maximizing Log likelihood is equivalent to minimizing the quadratic loss/error in the context of LR!

MLE Demonstration

MLE is standard & probabilistic technique

$$p(\mathbf{t}|\mathbf{x}, \mathbf{w}, \beta) = \prod_{n=1}^{N} \mathcal{N}\left(t_n | y(x_n, \mathbf{w}), \beta^{-1}\right).$$

$$= \prod_{n=1}^{N} \frac{1}{\sqrt{2\pi\beta^{-2}}} \exp\left\{-\frac{1}{2\beta^{-2}}\{y(x_n, \mathbf{w}) - t_n\}^2\right\}$$

Taking the log

$$\ln p(\mathbf{t}|\mathbf{x}, \mathbf{w}, \beta) = -\frac{\beta}{2}\sum_{n=1}^{N}\{y(x_n, \mathbf{w}) - t_n\}^2 + \frac{N}{2}\ln\beta - \frac{N}{2}\ln(2\pi).$$

It's called Log likelihood!

Maximizing Likelihood Learning

QA