CIS 678 - Machine Learning

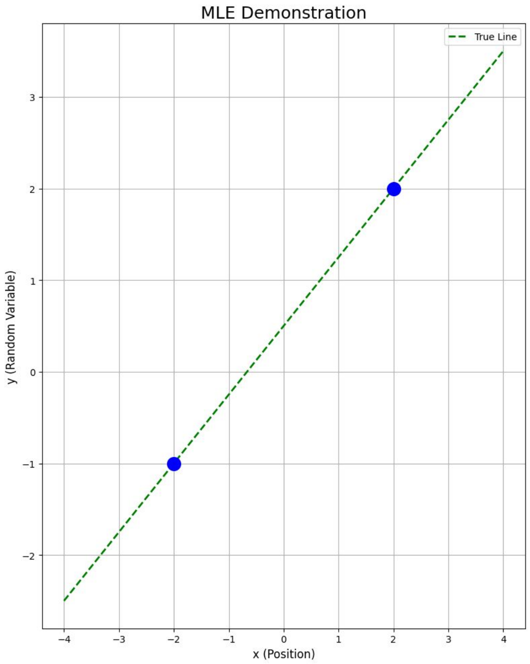
- Maximum Likelihood Learning

Linear Regression: Probabilistic Twin

Method of Least Squares

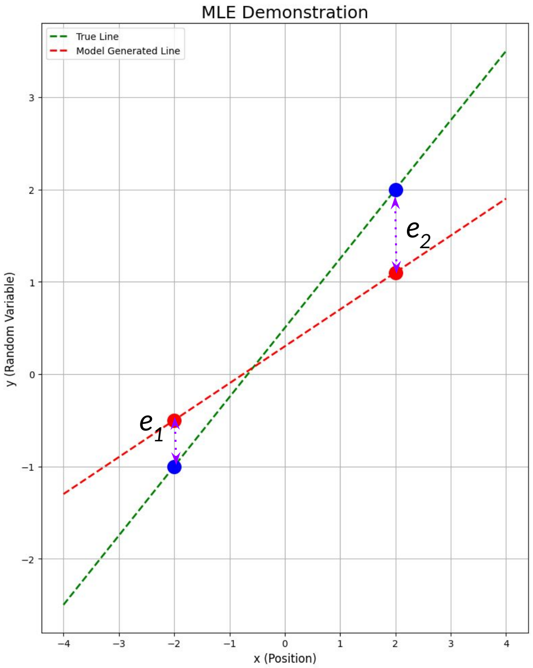
$$E(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{N}\{y(x_n, \mathbf{w}) - t_n\}^2$$

Linear Regression



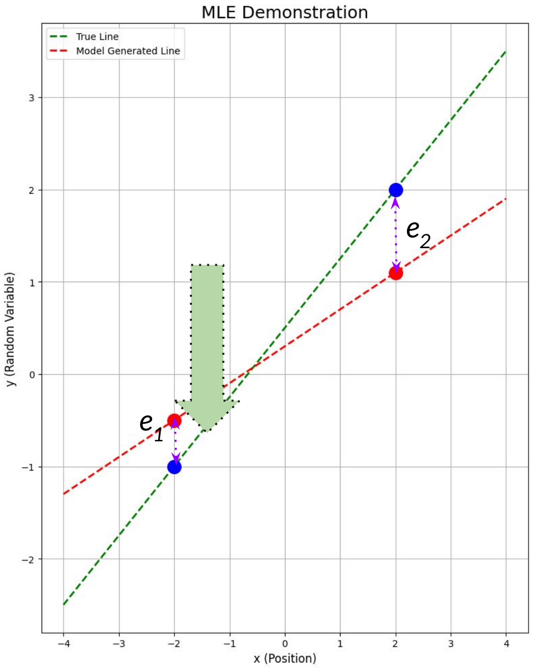
Least Squares Solution

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



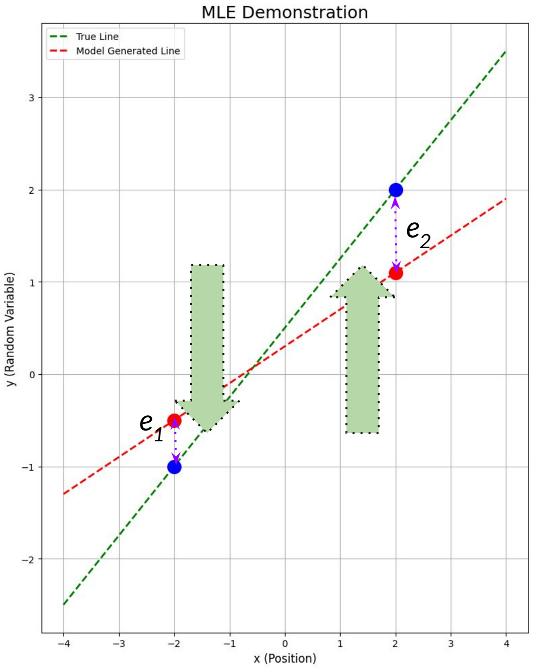
Least Squares Solution

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



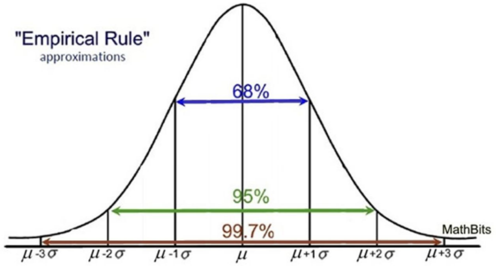
Least Squares Solution

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



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 $\mu \pm 3\sigma$



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

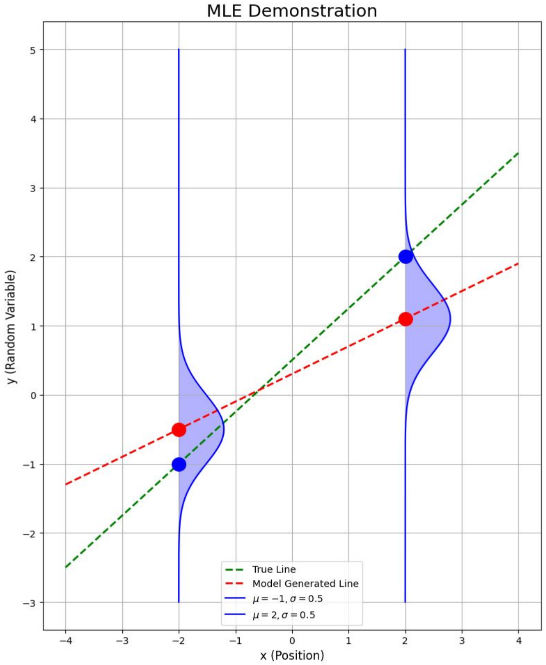
$\pi \approx 3.14159 \ldots$

$e \approx 2.71828 \ldots$

Probabilistic Twin

Probabilistic Formulation: Modeling Error Distribution

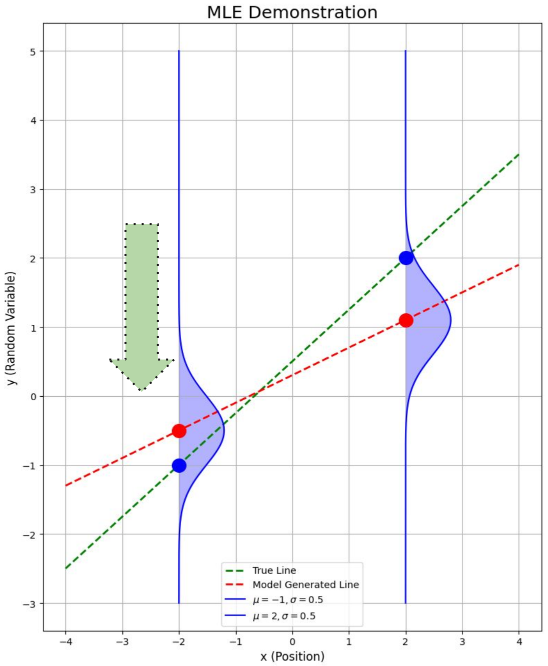
$$p(\mathbf{t}|\mathbf{x}, \mathbf{w}, \beta) = \prod_{n=1}^{N} \mathcal{N}\left(t_n | y(x_n, \mathbf{w}), \beta^{-1}\right).$$



Probabilistic Twin

Probabilistic Formulation: Modeling Error Distribution

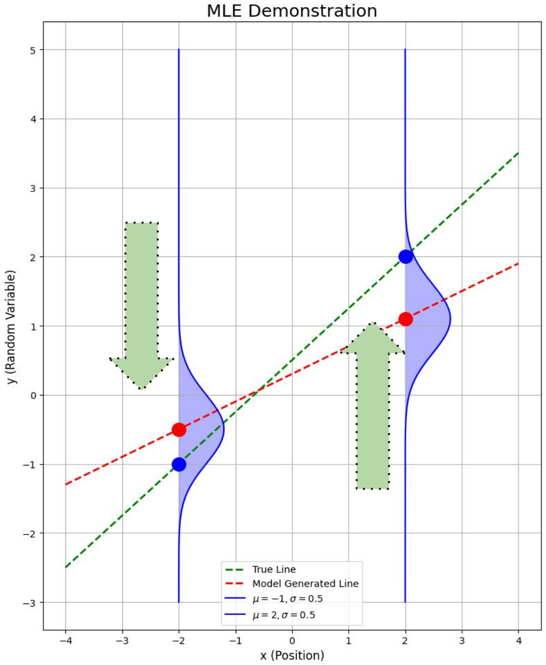
$$p(\mathbf{t}|\mathbf{x}, \mathbf{w}, \beta) = \prod_{n=1}^{N} \mathcal{N}\left(t_n | y(x_n, \mathbf{w}), \beta^{-1}\right).$$



Probabilistic Twin

Probabilistic Formulation: Modeling Error Distribution

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

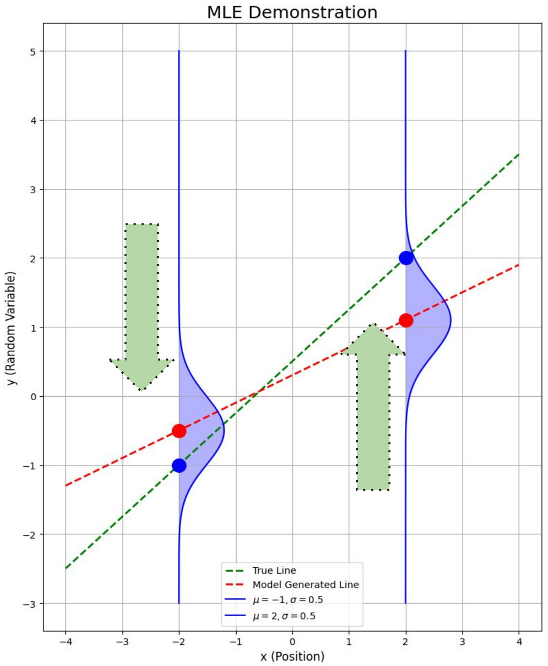


Probabilistic Twin

Probabilistic Formulation: Modeling Error Distribution

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

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

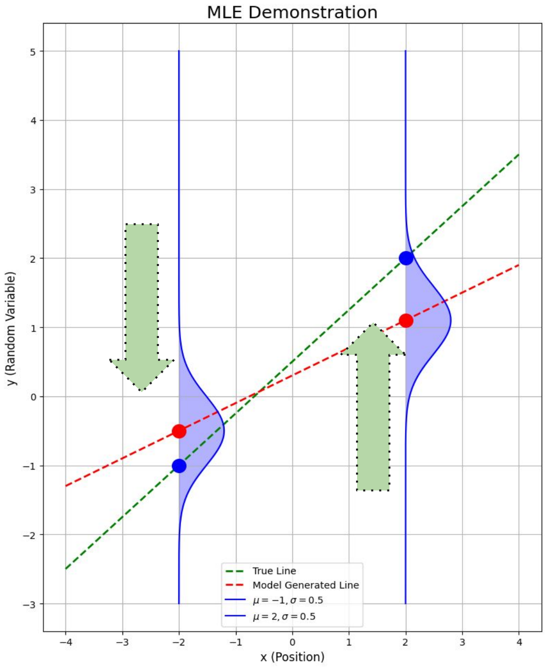


Probabilistic Twin

Probabilistic Formulation: Modeling Error Distribution

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

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



Probabilistic Twin

Probabilistic Formulation: Modeling Error Distribution

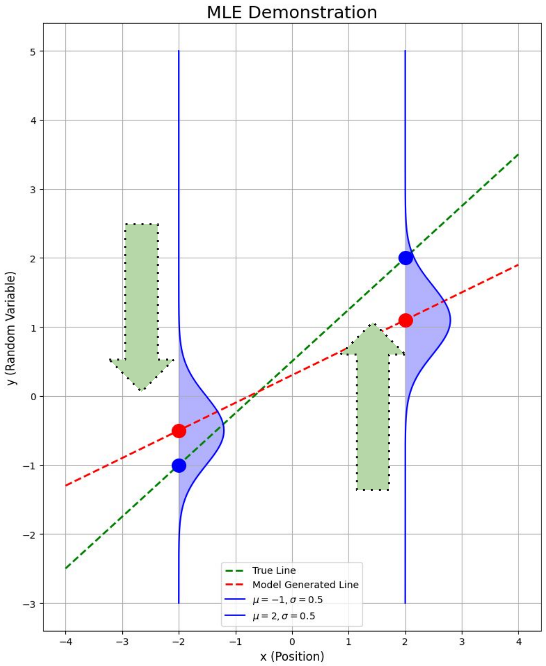
$$p(\mathbf{t}|\mathbf{x}, \mathbf{w}, \beta) = \prod_{n=1}^{N} \mathcal{N}\left(t_n | y(x_n, \mathbf{w}), \beta^{-1}\right).$$

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

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!



Probabilistic Twin

Probabilistic Formulation: Modeling Error Distribution

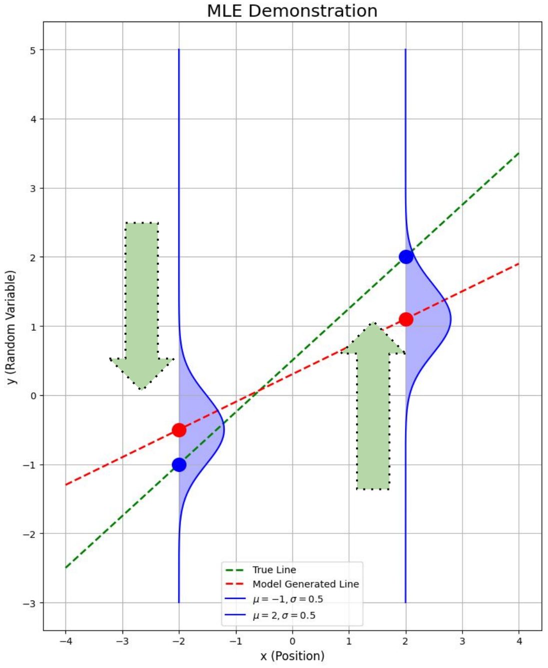
$$p(\mathbf{t}|\mathbf{x}, \mathbf{w}, \beta) = \prod_{n=1}^{N} \mathcal{N}\left(t_n | y(x_n, \mathbf{w}), \beta^{-1}\right).$$

<u>Taking the log</u>

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

Does it look familiar???

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



Probabilistic Twin

Probabilistic Formulation: Modeling Error Distribution

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

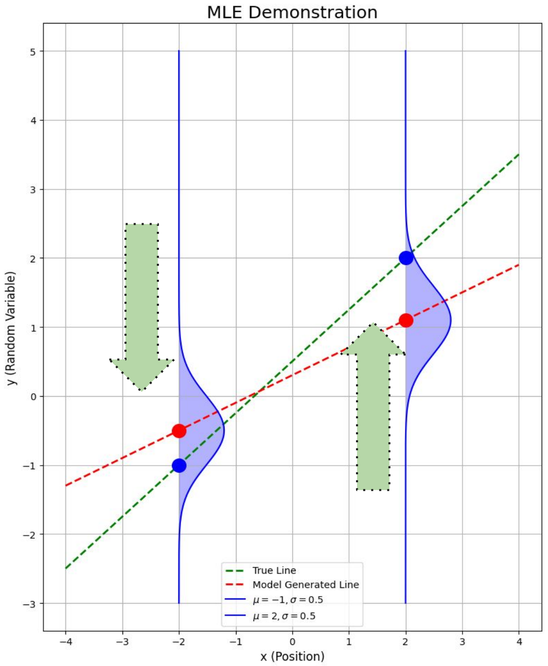
Taking the log

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

Does it look familiar???

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

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



MLE is standard & probabilistic technique

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!

Maximizing Likelihood Learning

$$\mathbf{w}^* = \arg\max_{W} \; p(\mathbf{t}|\mathbf{x}, \mathbf{w}, \beta^{-1})$$

MLE is standard & probabilistic technique

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!

Maximizing Likelihood Learning

Can be generalized for any problem given that we properly explain the distribution of the data

$$\mathbf{w}^* = \arg\max_{W} \; p(\mathbf{t}|\mathbf{x}, \mathbf{w}, \beta^{-1})$$

QA