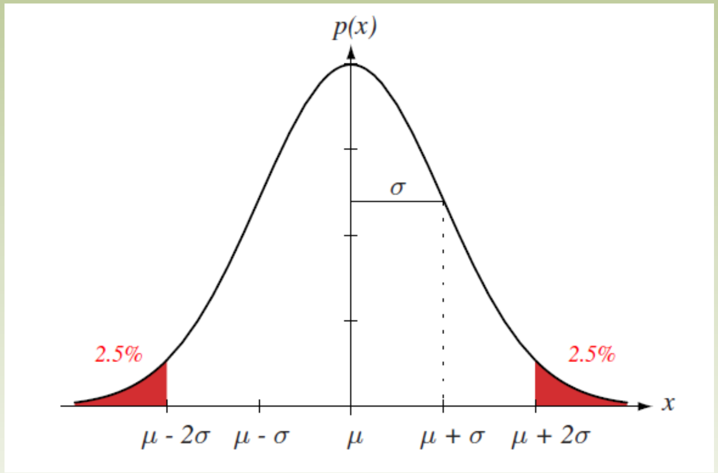
Linear Regression with Ordinary Least Squares

Univariate Normal (Gaussian) Distribution

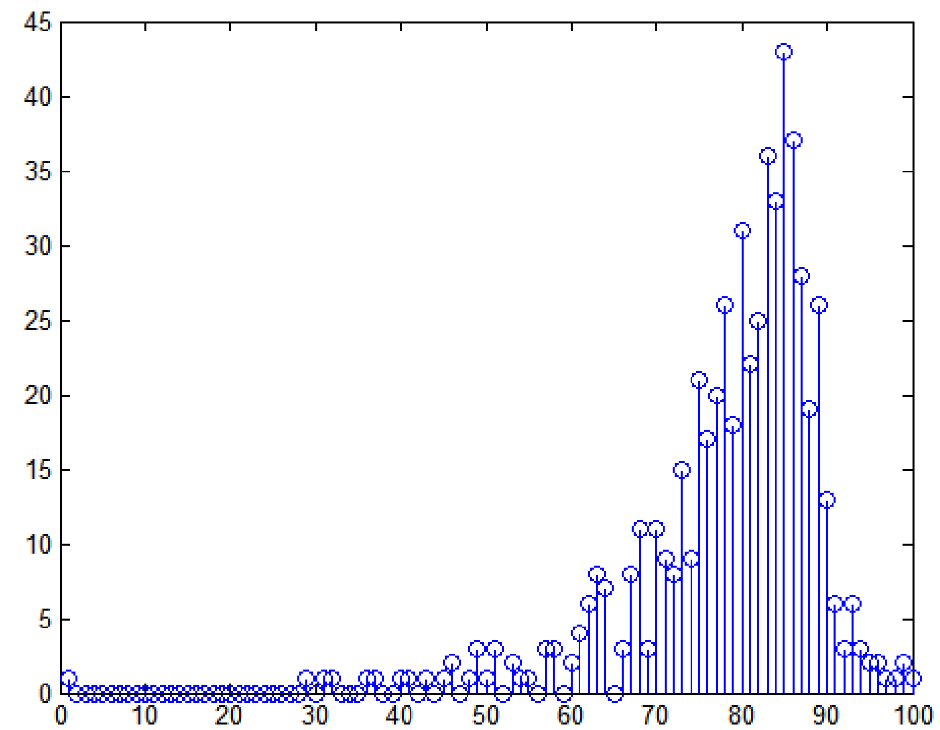
$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$



- The normal density is completely specified by the *mean* (μ) and the *variance* (σ^2). These two are its *sufficient statistics.*
- Short-hand notation: $p(x) \sim \mathcal{N}(\mu, \sigma^2)$

Linear Regression with OLS

- Error Histogram



Linear Regression with OLS

- Linear Regression is widely used in statistics and supervised machine learning
- Response is a linear function of the inputs: $y(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + \boldsymbol{\varepsilon}$
- $\boldsymbol{\varepsilon}$ is the residual error between model prediction and true response
  - $\varepsilon_i = y_i - \mathbf{w}^T\mathbf{x}_i$ → commonly assumed to be normally distributed with mean μ and variance σ^2
- Therefore, the density function $p(y|\mathbf{x}, \boldsymbol{\theta})$ is also Gaussian
  - $\boldsymbol{\mu}$ is a linear function of x → $\boldsymbol{\mu} = \mathbf{w}^T\mathbf{x}$
  - It is reasonable to assume the noise is fixed, i.e., $\sigma^2(\mathbf{x}) = \sigma^2$
  - Model parameters: $\boldsymbol{\theta} = (\boldsymbol{\mu}, \sigma^2) = (\mathbf{w}, \sigma^2)$

$$p(y|\mathbf{x}, \boldsymbol{\theta}) \sim \mathcal{N}(\boldsymbol{\mu}, \sigma^2) = \mathcal{N}(\mathbf{w}^T\mathbf{x}, \sigma^2)$$

Linear Regression with OLS

- Residual Sum of Squares (RSS)
  - Obtained from log likelihood
  - $RSS = \sum_{i=1}^{N}(y_i - \mathbf{w}^T\mathbf{x}_i)^2$
  - Also known as Sum of Squared Error (SSE)
  - Mean Squared Error (MSE) = SSE/N
- MLE minimizes RSS → Least Squares method
- Using MLE, the optimal weight vector can be computed.

$$\mathbf{w}_{OLS} = (\mathrm{X}^T\mathrm{X})^{-1}\mathrm{X}^T\mathbf{y}$$

- This is called the normal equation.
- This method is known as Ordinary Least Squares (OLS)

Linear Regression with OLS