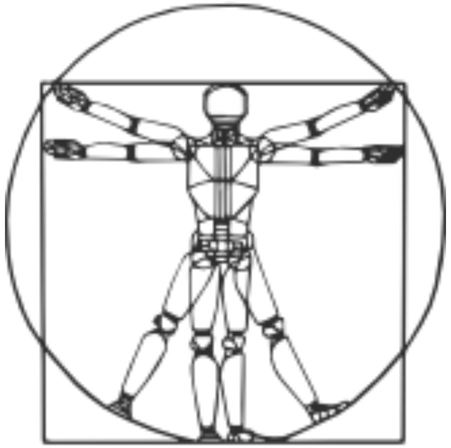
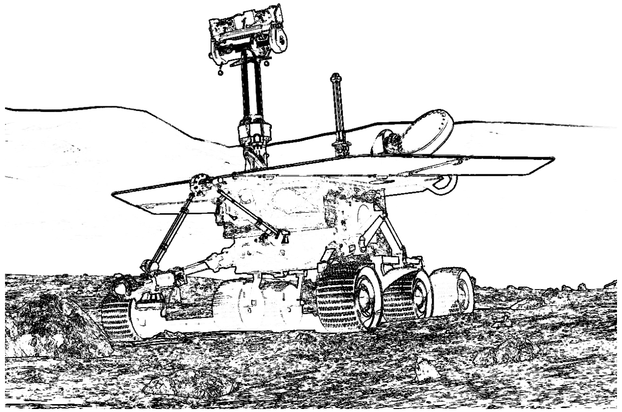




COMP417
Introduction to Robotics and Intelligent Systems

Lecture 8: Probabilistic Estimation 3 ways

Based on material by Florian Shkurti @ U of T

McGill

MRL Mobile Robotics Lab at McGill University

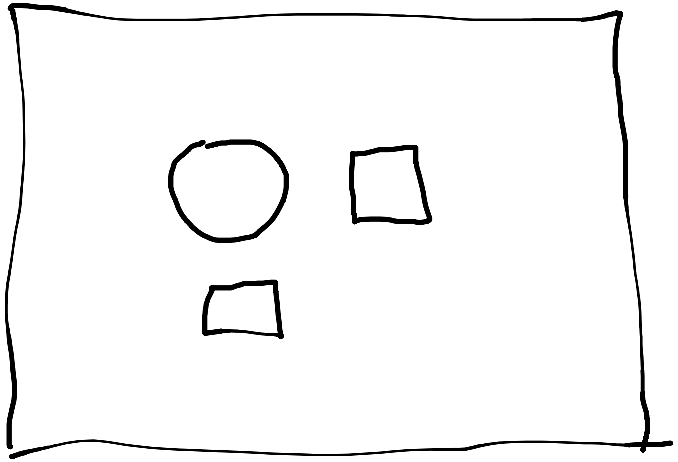
Estimating parameters of probability models

- In the occupancy grid mapping problem we wanted to compute $p(\mathbf{m}|\mathbf{z}_{1:t}, \mathbf{x}_{1:t})$ over all possible maps.

- While we previously treated the map as an occupancy grid, just 0,1 values, there is often more structure in the world that we'd like to capture:
  - The walls of a building are all straight
  - Corners appear only at right angles
  - Any rounded portions are roughly circular

- In the map on the right, a few numbers

  can very accurately describe large potions



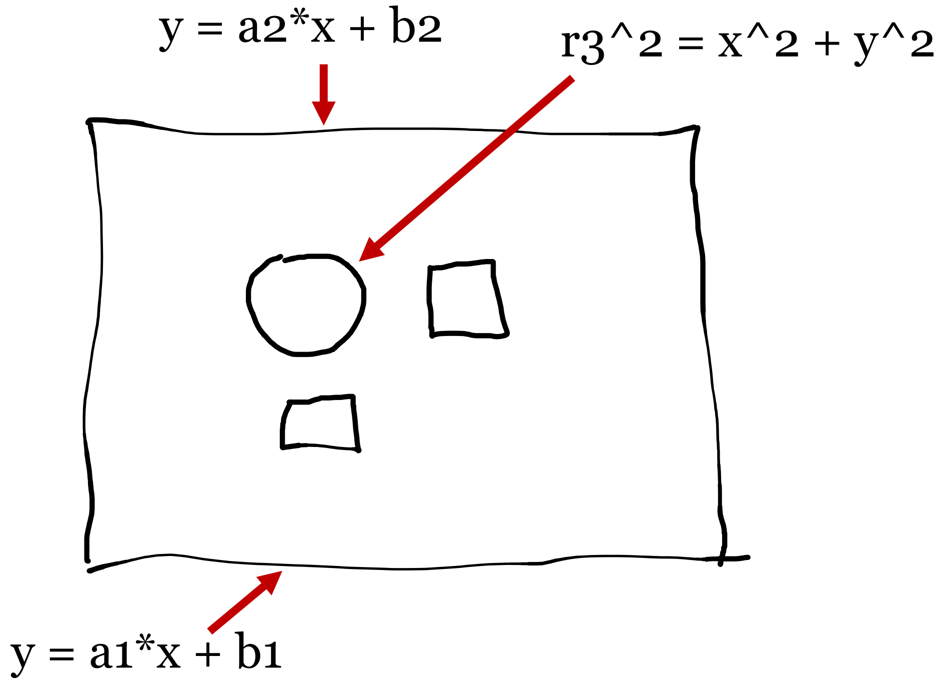
Estimating parameters of probability models

- In the occupancy grid mapping problem we wanted to compute $p(\mathbf{m}|\mathbf{z}_{1:t}, \mathbf{x}_{1:t})$ over all possible maps.

- While we previously treated the map as an occupancy grid, just 0,1 values, there is often more structure in the world that we'd like to capture:
  - The walls of a building are all straight
  - Corners appear only at right angles
  - Any rounded portions are roughly circular

- In the map on the right, a few numbers

  can very accurately describe large potions
  - a1, b1, a2, b2, r3… etc, until all basic shapes are covered.

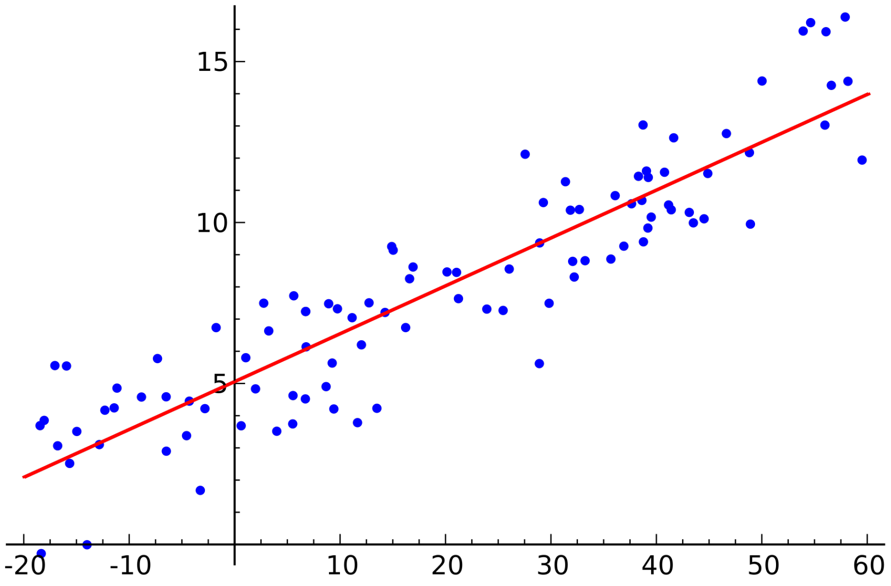


Estimating parameters of probability models

- Today we will look at several different classes of tools for finding the “parameters” of probability models that let us recover the map efficiently.

- There are typically three ways to work with this type of problems:
  1. Maximum Likelihood parameter estimation (MLE)
     - Least Squares
  2. Maximum A Posteriori (MAP) parameter estimation
  3. Bayesian parameter distribution estimation

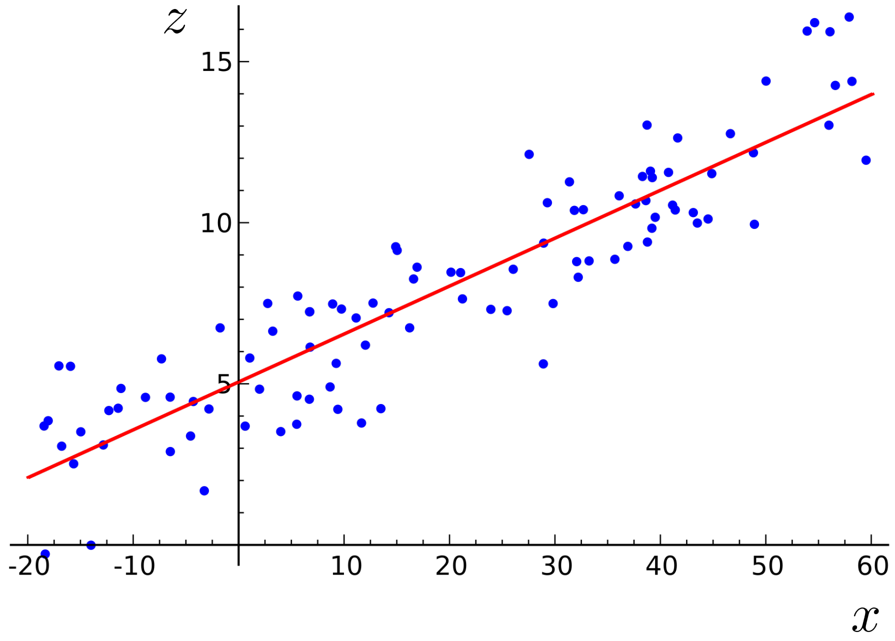
Least Squares Parameter Estimation



We are given data points $(\mathbf{x}_1, \mathbf{z}_1), \ldots, (\mathbf{x}_N, \mathbf{z}_N)$

We **think** that the data was generated by a parametric function $\mathbf{z} = \mathbf{h}(\boldsymbol{\theta}, \mathbf{x})$

Least Squares Parameter Estimation

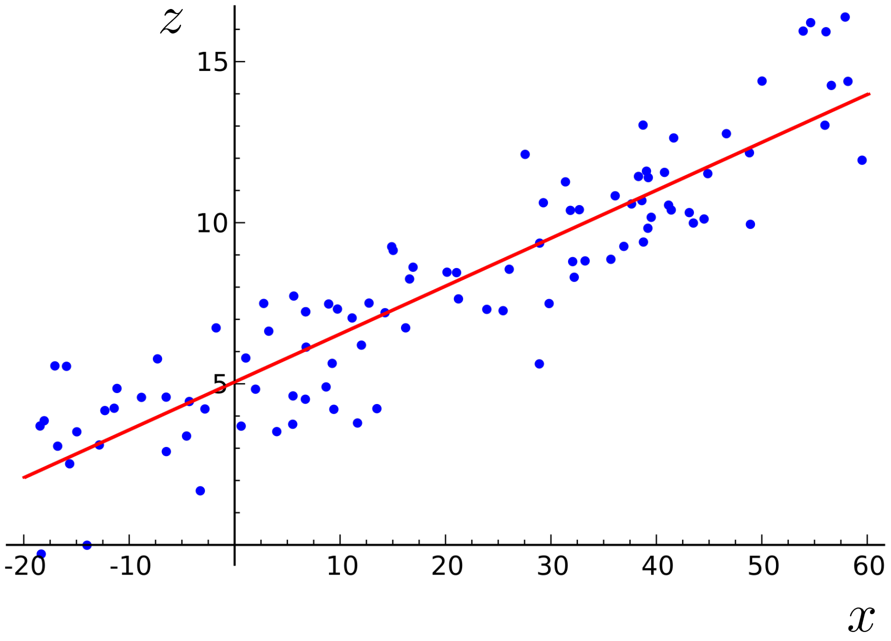


Example: we think that the 2D data was generated by a line $z = \theta_0 + \theta_1 x$ whose parameters we do not know, and was corrupted by noise.

We are given data points $(\mathbf{x}_1, \mathbf{z}_1), \ldots, (\mathbf{x}_N, \mathbf{z}_N)$

We **think** that the data was generated by a parametric function $\mathbf{z} = \mathbf{h}(\boldsymbol{\theta}, \mathbf{x})$

Least Squares Parameter Estimation



Example: we think that the 2D data was generated by a line $z = \theta_0 + \theta_1 x$ whose parameters we do not know.

We are given data points $(\mathbf{x}_1, \mathbf{z}_1), ..., (\mathbf{x}_N, \mathbf{z}_N)$

We **think** that the data was generated by a parametric function $\mathbf{z} = \mathbf{h}(\boldsymbol{\theta}, \mathbf{x})$

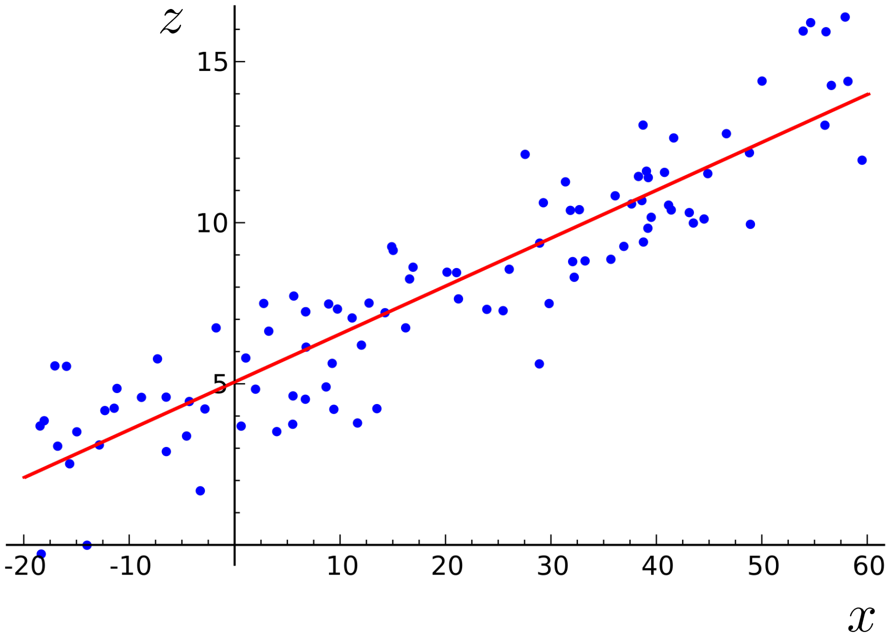
This parametric model will have a fitting error:

$$e(\boldsymbol{\theta}) = \sum_{i=1}^{N} ||\mathbf{z}_i - \mathbf{h}(\boldsymbol{\theta}, \mathbf{x}_i)||^2$$

The least-squares estimator is:

$$\boldsymbol{\theta}_{LS} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}}\ e(\boldsymbol{\theta})$$

Example #1: Linear Least Squares



Notice we stack θ_0 and θ_1 in a vector just to make the math clean.

We are given 2D data points $(x_1, z_1), \ldots, (x_N, z_N)$

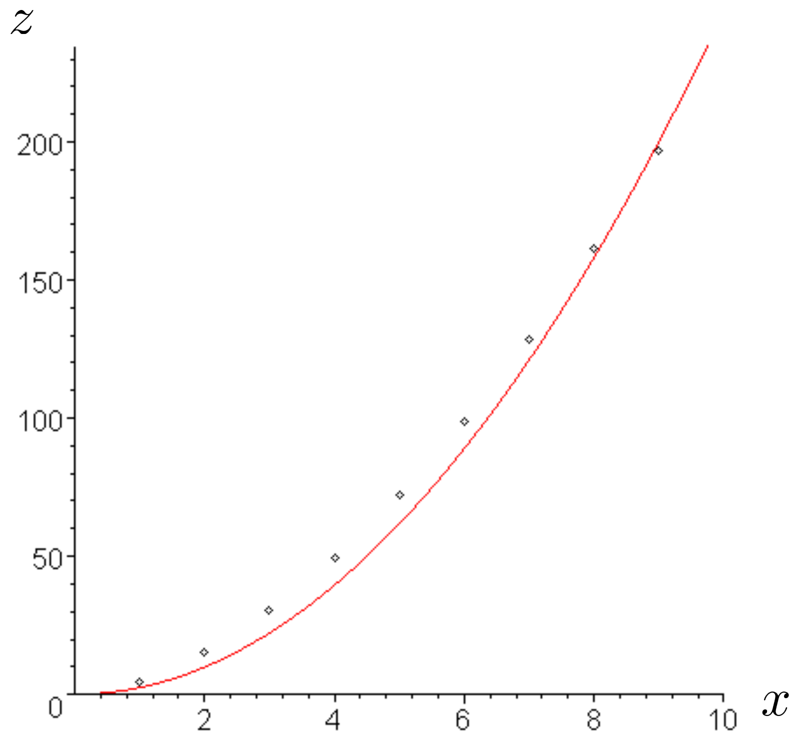
We **think** that the data was generated by a linear parametric function $z = h(\boldsymbol{\theta}, x) = [1 \quad x]\boldsymbol{\theta} = \theta_0 + \theta_1 x$

This parametric model will have a fitting error:

$$e(\theta_0, \theta_1) \quad = \quad \sum_{i=1}^{N} (z_i - \theta_0 - \theta_1 x_i)^2$$

Example: we think that the 2D data was generated by a line $z = \theta_0 + \theta_1 x$ whose parameters we do not know.

Example #2: Linear Least Squares



Example: we think that the 2D data was generated by a quadratic $z = \theta_0 + \theta_1 x + \theta_2 x^2$ whose parameters we do not know.

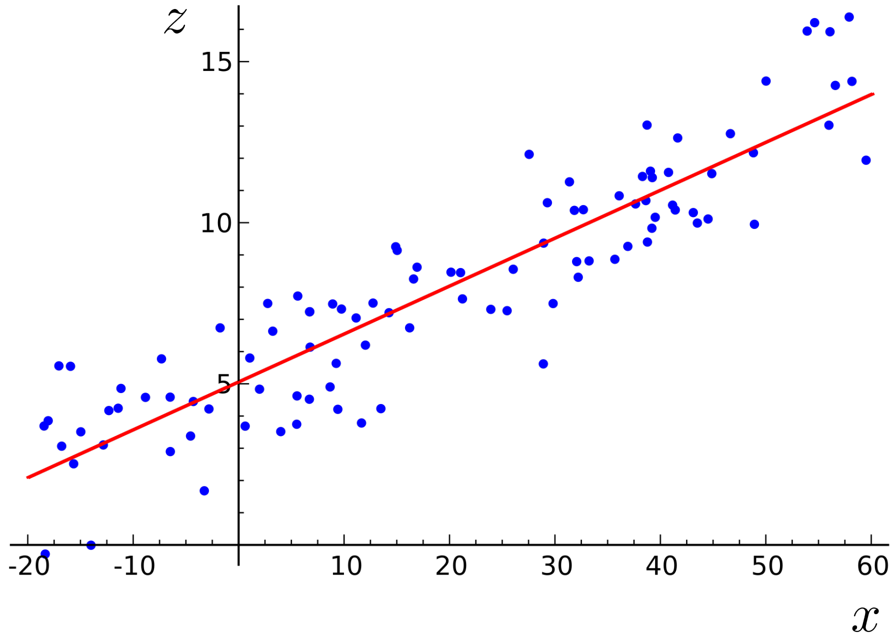
We are given 2D data points $(x_1, z_1), ..., (x_N, z_N)$

We **think** that the data was generated by a quadratic function $z = h(\boldsymbol{\theta}, x) = [1 \quad x \quad x^2]\boldsymbol{\theta} = \theta_0 + \theta_1 x + \theta_2 x^2$

This parametric model will have a fitting error:

$$e(\theta_0, \theta_1, \theta_2) \quad = \quad \sum_{i=1}^{N} (z_i - \theta_0 - \theta_1 x_i - \theta_2 x_i^2)^2$$

Generalize to all Linear Least Squares Parameter Estimations



Example: we think that the 2D data was generated by a line $z = \theta_0 + \theta_1 x$ whose parameters we do not know.

We are given data points $(\mathbf{x}_1, \mathbf{z}_1), ..., (\mathbf{x}_N, \mathbf{z}_N)$

We **think** that the data was generated by a linear parametric function $\mathbf{z} = \mathbf{h}(\boldsymbol{\theta}, \mathbf{x}) = \mathbf{H}_{\mathbf{x}}\boldsymbol{\theta}$
where $\mathbf{H}_{\mathbf{x}}$ is a matrix whose elements depend on $\mathbf{x}$

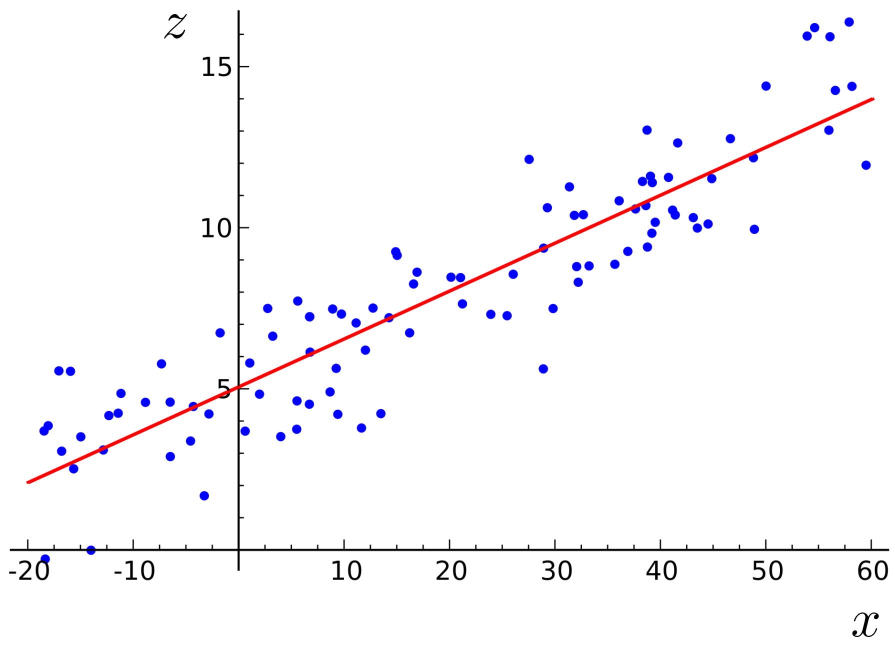
This parametric model will have a fitting error:

$$e(\boldsymbol{\theta}) = \sum_{i=1}^{N} ||\mathbf{z}_i - \mathbf{H}_{\mathbf{x}_i}\boldsymbol{\theta}||^2$$

The least-squares estimator is:

$$\boldsymbol{\theta}_{LS} = \underset{\boldsymbol{\theta}}{\text{argmin}}\ e(\boldsymbol{\theta})$$

Linear Least Squares Parameter Estimation



Example: we think that the 2D data was generated by a line $z = \theta_0 + \theta_1 x$ whose parameters we do not know.

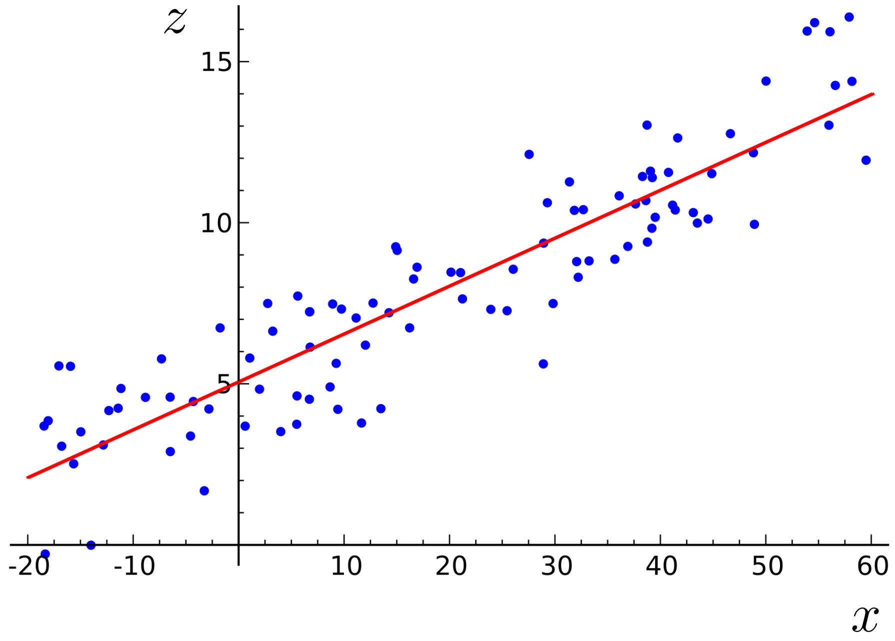
We are given data points $(\mathbf{x}_1, \mathbf{z}_1), ..., (\mathbf{x}_N, \mathbf{z}_N)$

We **think** that the data was generated by a linear parametric function $\mathbf{z} = \mathbf{h}(\boldsymbol{\theta}, \mathbf{x}) = \mathbf{H}_\mathbf{x}\boldsymbol{\theta}$

This parametric model will have a fitting error:

$$\begin{aligned} e(\boldsymbol{\theta}) &= \sum_{i=1}^{N} ||\mathbf{z}_i - \mathbf{H}_{\mathbf{x}_i}\boldsymbol{\theta}||^2 \\ &= \sum_{i=1}^{N} \mathbf{z}_i^T\mathbf{z}_i - 2\boldsymbol{\theta}^T\mathbf{H}_{\mathbf{x}_i}^T\mathbf{z}_i + \boldsymbol{\theta}^T\mathbf{H}_{\mathbf{x}_i}^T\mathbf{H}_{\mathbf{x}_i}\boldsymbol{\theta} \end{aligned}$$

Linear Least Squares Parameter Estimation



Example: we think that the 2D data was generated by a line $z = \theta_0 + \theta_1 x$ whose parameters we do not know.

We are given data points $(\mathbf{x}_1, \mathbf{z}_1), ..., (\mathbf{x}_N, \mathbf{z}_N)$

We **think** that the data was generated by a linear parametric function $\mathbf{z} = \mathbf{h}(\boldsymbol{\theta}, \mathbf{x}) = \mathbf{H}_{\mathbf{x}}\boldsymbol{\theta}$

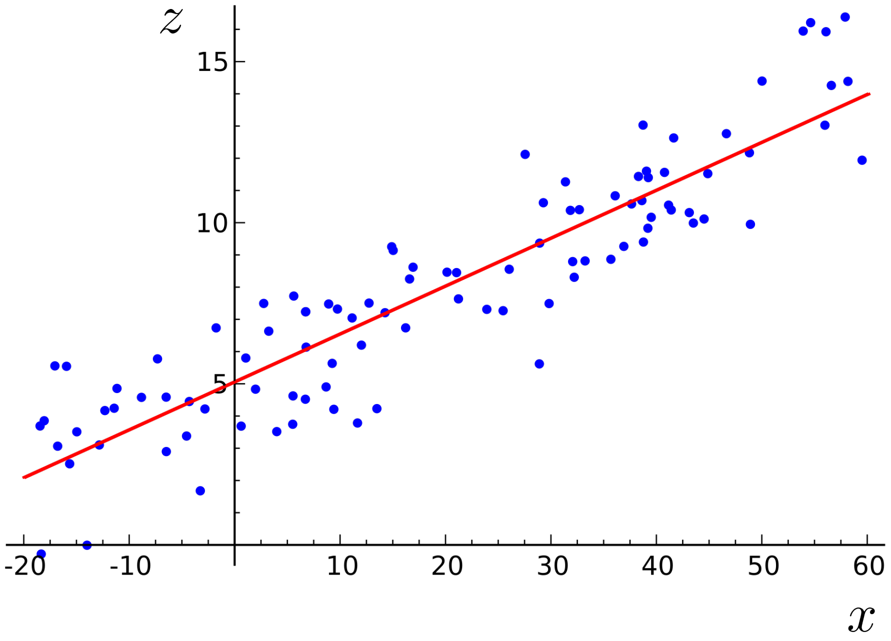
This parametric model will have a fitting error:

$$\begin{aligned} e(\boldsymbol{\theta}) &= \sum_{i=1}^{N} ||\mathbf{z}_i - \mathbf{H}_{\mathbf{x}_i}\boldsymbol{\theta}||^2 \\ &= \sum_{i=1}^{N} \mathbf{z}_i^T\mathbf{z}_i - 2\boldsymbol{\theta}^T\mathbf{H}_{\mathbf{x}_i}^T\mathbf{z}_i + \boldsymbol{\theta}^T\mathbf{H}_{\mathbf{x}_i}^T\mathbf{H}_{\mathbf{x}_i}\boldsymbol{\theta} \end{aligned}$$

The least-squares estimator minimizes the error:

$$\frac{\partial e(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{0} \Leftrightarrow -2\sum_{i=1}^{N} \mathbf{H}_{\mathbf{x}_i}^T\mathbf{z}_i + 2\mathbf{H}_{\mathbf{x}_i}^T\mathbf{H}_{\mathbf{x}_i}\boldsymbol{\theta} = \mathbf{0} \Leftrightarrow \left[\sum_{i=1}^{N} \mathbf{H}_{\mathbf{x}_i}^T\mathbf{H}_{\mathbf{x}_i}\right]\boldsymbol{\theta} = \sum_{i=1}^{N} \mathbf{H}_{\mathbf{x}_i}^T\mathbf{z}_i$$

Linear Least Squares Parameter Estimation



Example: we think that the 2D data was generated by a line $z = \theta_0 + \theta_1 x$ whose parameters we do not know.

We are given data points $(\mathbf{x}_1, \mathbf{z}_1), ..., (\mathbf{x}_N, \mathbf{z}_N)$

We **think** that the data was generated by a linear parametric function $\mathbf{z} = \mathbf{h}(\boldsymbol{\theta}, \mathbf{x}) = \mathbf{H}_{\mathbf{x}}\boldsymbol{\theta}$

This parametric model will have a fitting error:

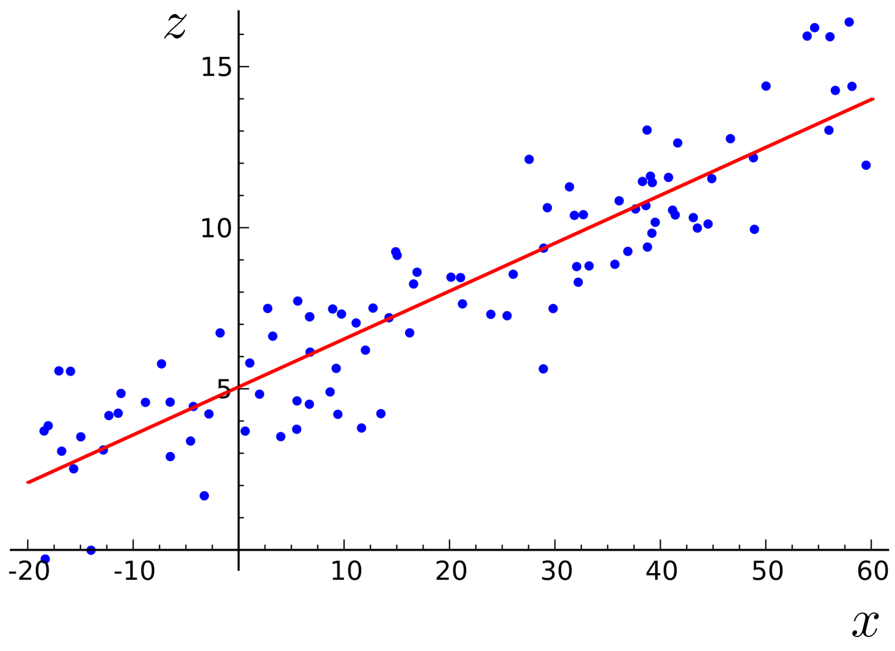
$$\begin{aligned} e(\boldsymbol{\theta}) &= \sum_{i=1}^{N} ||\mathbf{z}_i - \mathbf{H}_{\mathbf{x}_i}\boldsymbol{\theta}||^2 \\ &= \sum_{i=1}^{N} \mathbf{z}_i^T \mathbf{z}_i - 2\boldsymbol{\theta}^T \mathbf{H}_{\mathbf{x}_i}^T \mathbf{z}_i + \boldsymbol{\theta}^T \mathbf{H}_{\mathbf{x}_i}^T \mathbf{H}_{\mathbf{x}_i}\boldsymbol{\theta} \end{aligned}$$

The least-squares estimator minimizes the error:

$$\frac{\partial e(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{0} \Leftrightarrow -2\sum_{i=1}^{N} \mathbf{H}_{\mathbf{x}_i}^T \mathbf{z}_i + 2\mathbf{H}_{\mathbf{x}_i}^T \mathbf{H}_{\mathbf{x}_i}\boldsymbol{\theta} = \mathbf{0} \Leftrightarrow \left[\sum_{i=1}^{N} \mathbf{H}_{\mathbf{x}_i}^T \mathbf{H}_{\mathbf{x}_i}\right]\boldsymbol{\theta} = \sum_{i=1}^{N} \mathbf{H}_{\mathbf{x}_i}^T \mathbf{z}_i$$

$$\boldsymbol{\theta}_{LS} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}}\ e(\boldsymbol{\theta}) \Leftrightarrow \left[\sum_{i=1}^{N} \mathbf{H}_{\mathbf{x}_i}^T \mathbf{H}_{\mathbf{x}_i}\right]\boldsymbol{\theta}_{LS} = \sum_{i=1}^{N} \mathbf{H}_{\mathbf{x}_i}^T \mathbf{z}_i$$

Solution #1: Linear Least Squares



Example: we think that the 2D data was generated by a line $z = \theta_0 + \theta_1 x$ whose parameters we do not know.

We are given 2D data points $(x_1, z_1), \ldots, (x_N, z_N)$

We **think** that the data was generated by a linear parametric function $z = h(\boldsymbol{\theta}, x) = [1 \quad x]\boldsymbol{\theta} = \theta_0 + \theta_1 x$

This parametric model will have a fitting error:

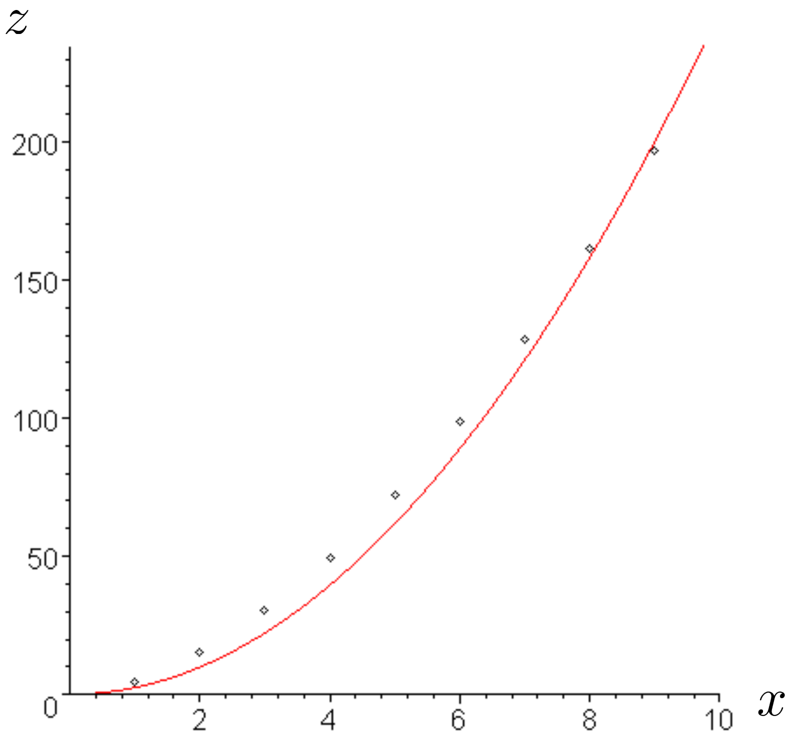
$$e(\theta_0, \theta_1) \quad = \quad \sum_{i=1}^{N} (z_i - \theta_0 - \theta_1 x_i)^2$$

The least-squares estimator minimizes the error:

$$\boldsymbol{\theta}_{LS} = \underset{\theta_0, \theta_1}{\text{argmin}}\ e(\theta_0, \theta_1) \Leftrightarrow \left[\sum_{i=1}^{N} \begin{bmatrix} 1 \\ x_i \end{bmatrix} \begin{bmatrix} 1 & x_i \end{bmatrix} \right] \boldsymbol{\theta}_{LS} = \sum_{i=1}^{N} \begin{bmatrix} 1 \\ x_i \end{bmatrix} z_i$$

Which is a linear system of 2 equations. If we have at least two data points we can solve for $\boldsymbol{\theta}_{LS}$ to define the line.

Solution #2: Linear Least Squares



Example: we think that the 2D data was generated by a quadratic $z = \theta_0 + \theta_1 x + \theta_2 x^2$ whose parameters we do not know.

We are given 2D data points $(x_1, z_1), ..., (x_N, z_N)$

We **think** that the data was generated by a quadratic function $z = h(\boldsymbol{\theta}, x) = [1 \quad x \quad x^2]\boldsymbol{\theta} = \theta_0 + \theta_1 x + \theta_2 x^2$

This parametric model will have a fitting error:

$$e(\theta_0, \theta_1, \theta_2) \quad = \quad \sum_{i=1}^{N} (z_i - \theta_0 - \theta_1 x_i - \theta_2 x_i^2)^2$$

The least-squares estimator minimizes the error:

$$\boldsymbol{\theta}_{LS} = \underset{\theta_0, \theta_1, \theta_2}{\text{argmin}}\ e(\theta_0, \theta_1, \theta_2) \Leftrightarrow \left[\sum_{i=1}^{N} \begin{bmatrix} 1 \\ x_i \\ x_i^2 \end{bmatrix} [1 \quad x_i \quad x_i^2] \right] \boldsymbol{\theta}_{LS} = \sum_{i=1}^{N} \begin{bmatrix} 1 \\ x_i \\ x_i^2 \end{bmatrix} z_i$$

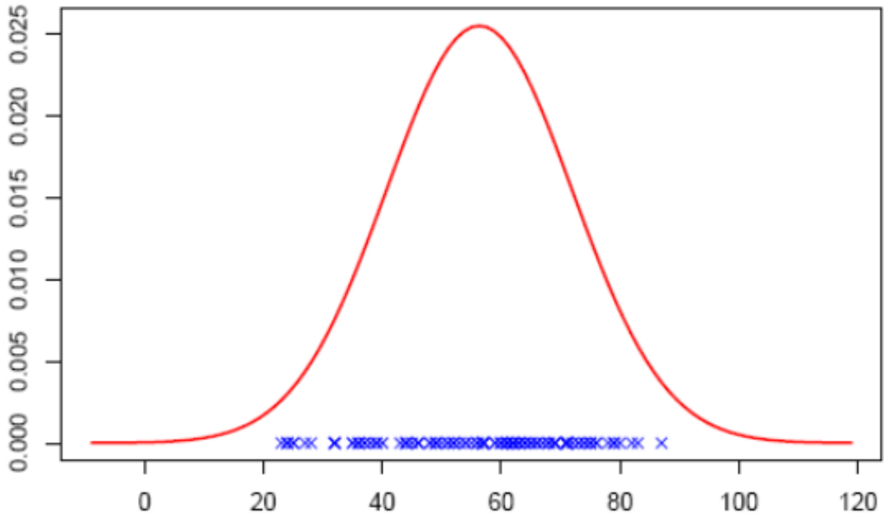
Which is a linear system of 3 equations. If we have at least three data points we can solve for $\boldsymbol{\theta}_{LS}$ to define the quadratic.

Estimating parameters of probability models

- Today we will look at several different classes of tools for finding the "parameters" of probability models that let us recover the map efficiently.

- There are typically three ways to work with this type of problems:
  1. Maximum Likelihood parameter estimation (MLE)
     - Least Squares
  2. Maximum A Posteriori (MAP) parameter estimation
  3. Bayesian parameter distribution estimation

Maximum Likelihood Parameter Estimation



We are given data points $\mathbf{d}_{1:N} = \mathbf{d}_1, ..., \mathbf{d}_N$

We **think** the data has been generated from a probability distribution $p(\mathbf{d}_{1:N}|\boldsymbol{\theta})$

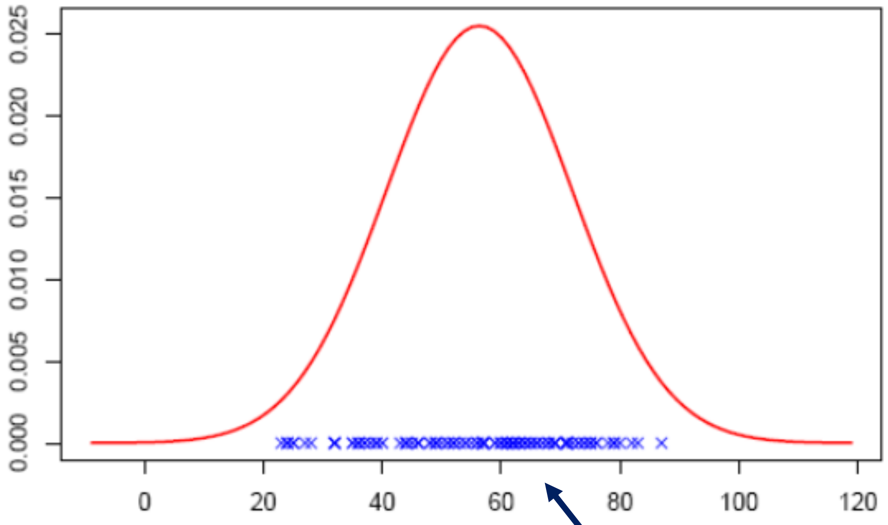
We want to find the parameter of the model that maximizes the likelihood function of the data

$$L(\boldsymbol{\theta}) = p(\mathbf{d}_{1:N}|\boldsymbol{\theta})$$

which is a function of theta, **not** a probability distribution.

$$\boldsymbol{\theta}_{MLE} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, p(\mathbf{d}_{1:N}|\boldsymbol{\theta})$$

Maximum Likelihood Parameter Estimation



Data points

$$\boldsymbol{\theta}_{MLE} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\ p(\mathbf{d}_{1:N}|\boldsymbol{\theta})$$

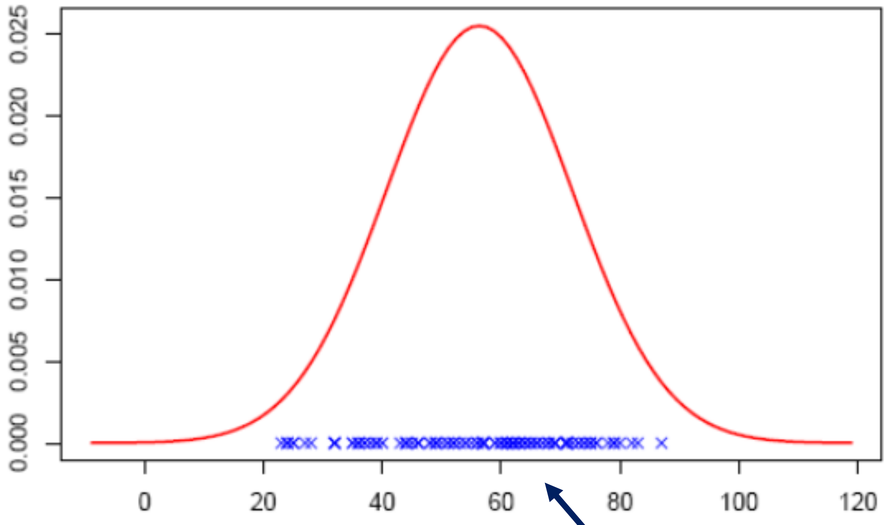
Find the parameters of the model that maximize the likelihood function of the data

$$L(\boldsymbol{\theta}) = p(\mathbf{d}_{1:N}|\boldsymbol{\theta})$$

which is a function of theta, **not** a probability distribution.

Example: assume we know that 1D data points were generated independently from a Gaussian distribution $\mathcal{N}(\mu, \sigma^2)$, but we don't know the mean and variance. The likelihood function of the data is

Maximum Likelihood Parameter Estimation



Data points

$$\boldsymbol{\theta}_{MLE} = \underset{\boldsymbol{\theta}}{\text{argmax}}\; p(\mathbf{d}_{1:N}|\boldsymbol{\theta})$$

Find the parameters of the model that maximize the likelihood function of the data

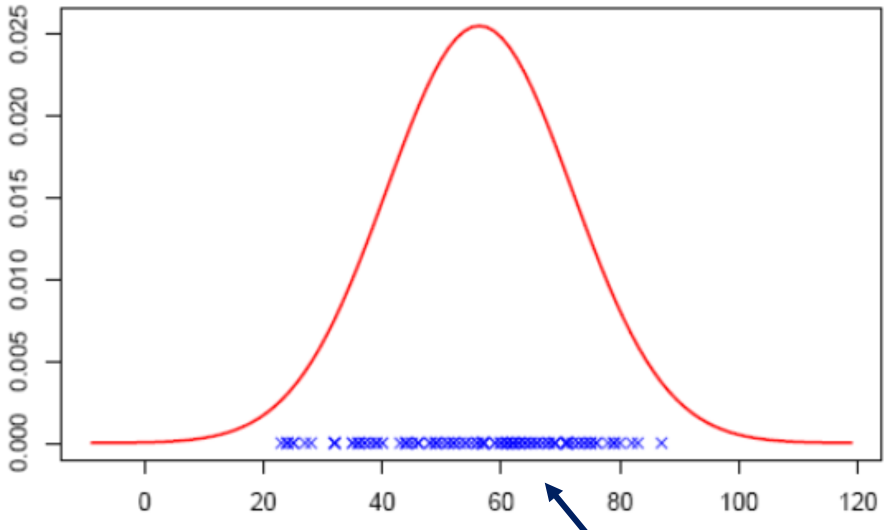
$$L(\boldsymbol{\theta}) = p(\mathbf{d}_{1:N}|\boldsymbol{\theta})$$

which is a function of theta, **not** a probability distribution.

Example: assume we know that 1D data points were generated <u>independently</u> from a Gaussian distribution $\mathcal{N}(\mu, \sigma^2)$, but we don't know the mean and variance. The likelihood function of the data is

$$L(\mu, \sigma) = p(\mathbf{d}_{1:N}|\mu, \sigma) = \prod_{i=1}^{N} p(d_i|\mu, \sigma) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma}\exp(-0.5(d_i - \mu)^2/\sigma^2)$$

Maximum Likelihood Parameter Estimation



Data points

$$\boldsymbol{\theta}_{MLE} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, p(\mathbf{d}_{1:N}|\boldsymbol{\theta})$$

Find the parameters of the model that maximize the likelihood function of the data

$$L(\boldsymbol{\theta}) = p(\mathbf{d}_{1:N}|\boldsymbol{\theta})$$

which is a function of theta, **not** a probability distribution.

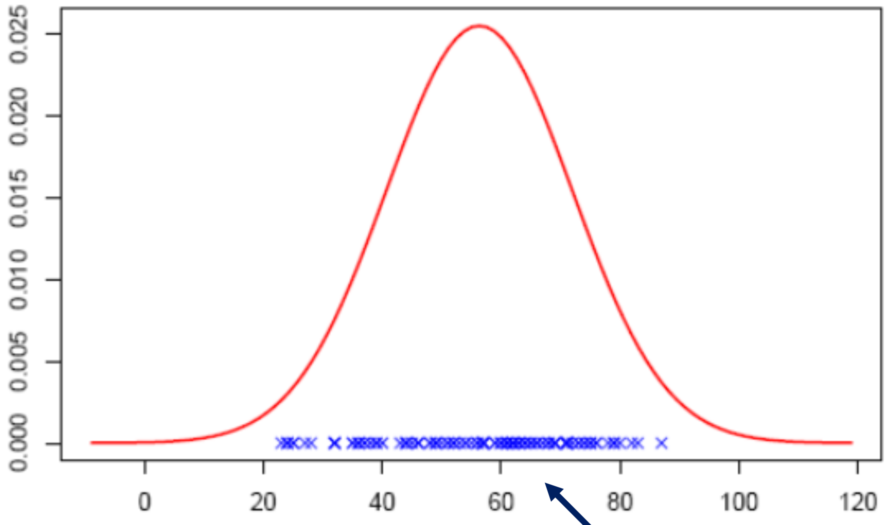
Example: assume we know that 1D data points were generated independently from a Gaussian distribution $\mathcal{N}(\mu, \sigma^2)$, but we don't know the mean and variance. The likelihood function of the data is

$$L(\mu, \sigma) = p(\mathbf{d}_{1:N}|\mu, \sigma) = \prod_{i=1}^{N} p(d_i|\mu, \sigma) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma} \exp(-0.5(d_i - \mu)^2/\sigma^2)$$

And the maximum-likelihood parameter estimates are

$$(\mu, \sigma)_{MLE} = \underset{\mu,\sigma}{\operatorname{argmax}}\, p(\mathbf{d}_{1:N}|\mu, \sigma) = \underset{\mu,\sigma}{\operatorname{argmax}}\, \log p(\mathbf{d}_{1:N}|\mu, \sigma) = \underset{\mu,\sigma}{\operatorname{argmax}} \sum_{i=1}^{N} \log p(d_i|\mu, \sigma)$$

Maximum Likelihood Parameter Estimation



Data points

$$\boldsymbol{\theta}_{MLE} = \underset{\boldsymbol{\theta}}{\text{argmax}}\; p(\mathbf{d}_{1:N}|\boldsymbol{\theta})$$

Find the parameters of the model that maximize the likelihood function of the data

$$L(\boldsymbol{\theta}) = p(\mathbf{d}_{1:N}|\boldsymbol{\theta})$$

which is a function of theta, **not** a probability distribution.

Example: assume we know that 1D data points were generated independently from a Gaussian distribution $\mathcal{N}(\mu, \sigma^2)$, but we don't know the mean and variance. The likelihood function of the data is

$$L(\mu, \sigma) = p(\mathbf{d}_{1:N}|\mu, \sigma) = \prod_{i=1}^{N} p(d_i|\mu, \sigma) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma} \exp(-0.5(d_i - \mu)^2/\sigma^2)$$

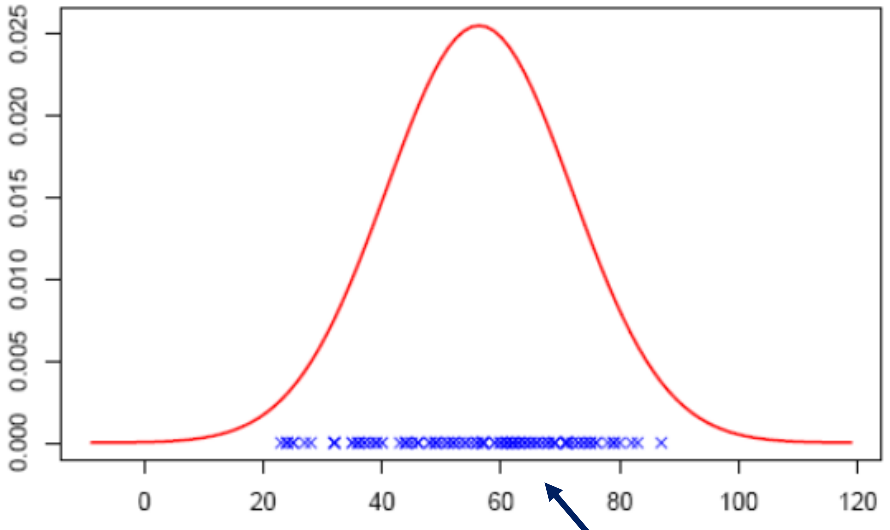
And the maximum-likelihood parameter estimates are

$$(\mu, \sigma)_{MLE} = \underset{\mu,\sigma}{\text{argmax}} \sum_{i=1}^{N} \log p(d_i|\mu, \sigma) = \underset{\mu,\sigma}{\text{argmax}} \left[-N\log(\sqrt{2\pi}\sigma) - \frac{1}{2\sigma^2} \sum_{i=1}^{N} (d_i - \mu)^2 \right]$$

Maximum Likelihood Parameter Estimation

Data points

$$\boldsymbol{\theta}_{MLE} = \underset{\boldsymbol{\theta}}{\text{argmax}}\; p(\mathbf{d}_{1:N}|\boldsymbol{\theta})$$



Find the parameters of the model that maximize the likelihood function of the data

$$L(\boldsymbol{\theta}) = p(\mathbf{d}_{1:N}|\boldsymbol{\theta})$$

which is a function of theta, **not** a probability distribution.

Example: assume we know that 1D data points were generated independently from a Gaussian distribution $\mathcal{N}(\mu, \sigma^2)$, but we don't know the mean and variance. The likelihood function of the data is

$$L(\mu, \sigma) = p(\mathbf{d}_{1:N}|\mu, \sigma) = \prod_{i=1}^{N} p(d_i|\mu, \sigma) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma} \exp(-0.5(d_i - \mu)^2/\sigma^2)$$

And the maximum-likelihood parameter estimates are

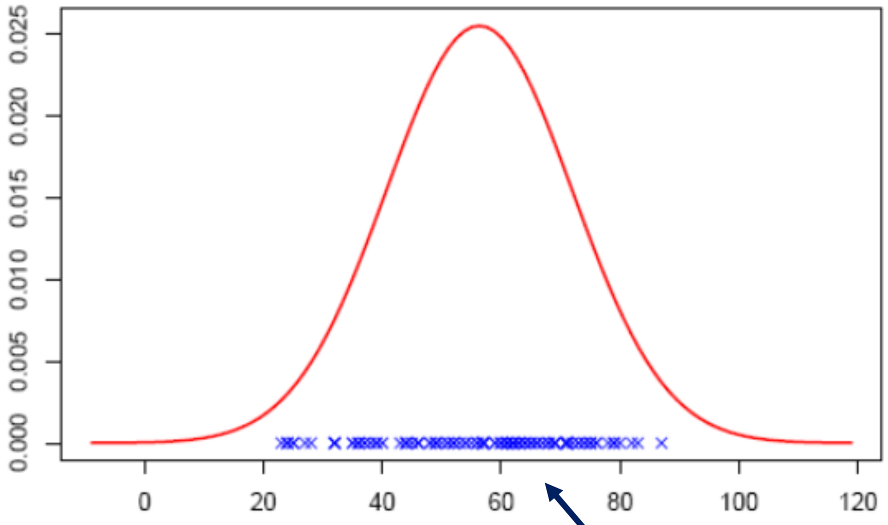
$$(\mu, \sigma)_{MLE} = \underset{\mu,\sigma}{\text{argmax}} \sum_{i=1}^{N} \log p(d_i|\mu, \sigma) = \underset{\mu,\sigma}{\text{argmax}} \left[-N\log(\sqrt{2\pi}\sigma) - \frac{1}{2\sigma^2} \sum_{i=1}^{N} (d_i - \mu)^2 \right]$$

Set partial derivatives w.r.t. μ and σ to zero

$$\mu_{MLE} = \sum_{i=1}^{N} d_i/N$$

$$\sigma^2_{MLE} = \frac{1}{N} \sum_{i=1}^{N} (d_i - \mu_{MLE})^2$$

Least Squares as Maximum Likelihood



$$\boldsymbol{\theta}_{MLE} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, p(\mathbf{d}_{1:N}|\boldsymbol{\theta})$$

Find the parameters of the model that maximize the likelihood function of the data

$$L(\boldsymbol{\theta}) = p(\mathbf{d}_{1:N}|\boldsymbol{\theta})$$

which is a function of theta, **not** a probability distribution.

Example: assume we know that 1D data points were generated independently from a Gaussian distribution $\mathcal{N}(\mu, \sigma^2)$, but we don't know the mean and variance. The likelihood function of the data is

$$L(\mu,\sigma) = p(\mathbf{d}_{1:N}|\mu,\sigma) = \prod_{i=1}^{N} p(d_i|\mu,\sigma) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma}\exp(-0.5(d_i-\mu)^2/\sigma^2)$$

And the maximum-likelihood parameter estimates are

$$(\mu,\sigma)_{MLE} = \underset{\mu,\sigma}{\operatorname{argmax}} \sum_{i=1}^{N} \log p(d_i|\mu,\sigma) = \underset{\mu,\sigma}{\operatorname{argmax}} \left[-N\log(\sqrt{2\pi}\sigma) - \frac{1}{2\sigma^2}\sum_{i=1}^{N}(d_i-\mu)^2\right]$$

Least squares estimation occurs in maximum likelihood with Gaussian models of data

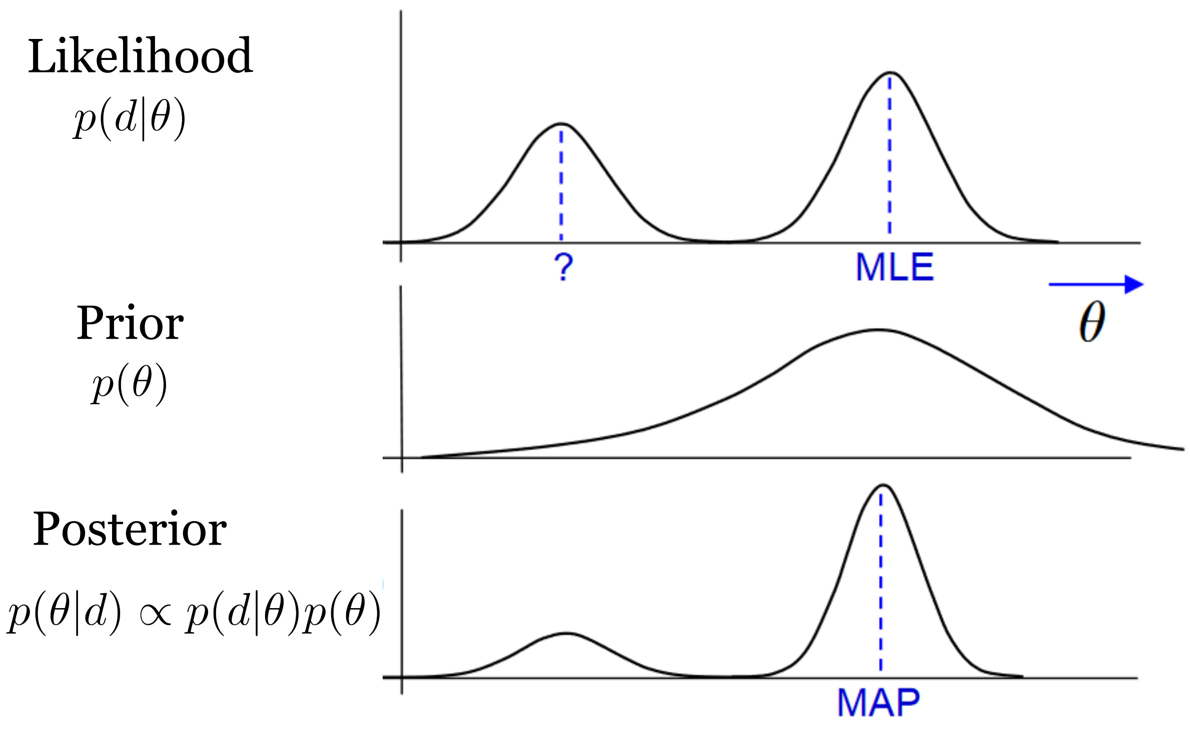
Estimating parameters of probability models

- In the occupancy grid mapping problem we wanted to compute $p(\mathbf{m}|\mathbf{z}_{1:t}, \mathbf{x}_{1:t})$ over all possible maps.

- We can see this problem as a specific instance within a category of problems where we are given data (observations) and we want to “explain” or fit the data using a parametric function.

- There are typically three ways to work with this type of problems:
  1. Maximum Likelihood parameter estimation (MLE)
     - Least Squares
  2. Maximum A Posteriori (MAP) parameter estimation
  3. Bayesian parameter distribution estimation

Maximum A Posteriori Parameter Estimation



$$\begin{aligned}\boldsymbol{\theta}_{MAP} &= \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, p(\boldsymbol{\theta}|\mathbf{d}_{1:N}) \\ &= \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \left[\frac{p(\mathbf{d}_{1:N}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathbf{d}_{1:N})}\right]\end{aligned}$$

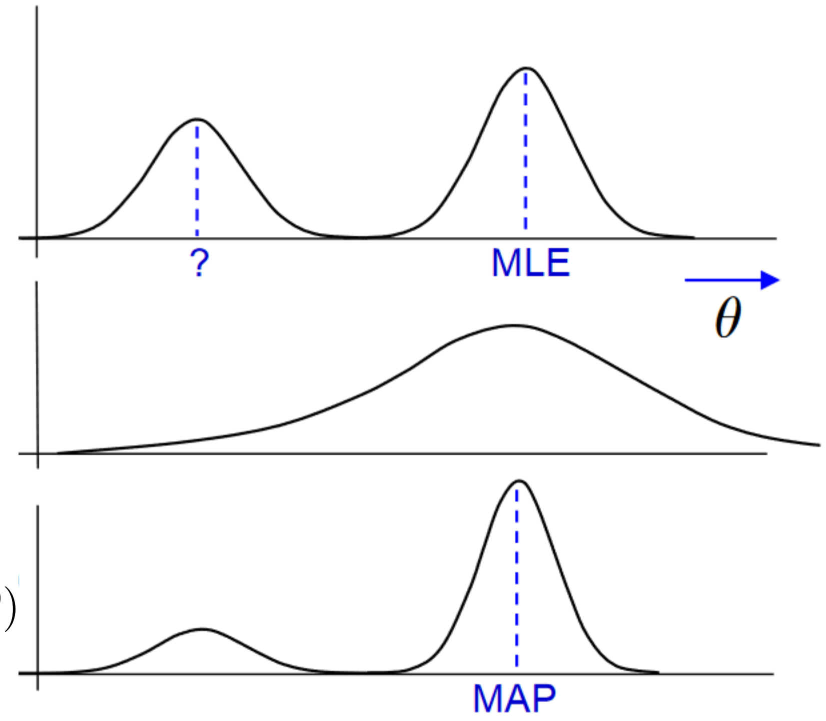
Note the denominator does not depend on the parameters.

Maximum A Posteriori Parameter Estimation

Likelihood
$p(d|\theta)$

Prior
$p(\theta)$

Posterior
$p(\theta|d) \propto p(d|\theta)p(\theta)$



$$
\begin{aligned}
\boldsymbol{\theta}_{MAP} &= \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, p(\boldsymbol{\theta}|\mathbf{d}_{1:N}) \\
&= \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \left[\frac{p(\mathbf{d}_{1:N}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathbf{d}_{1:N})}\right] \\
&= \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \left[p(\mathbf{d}_{1:N}|\boldsymbol{\theta})p(\boldsymbol{\theta})\right]
\end{aligned}
$$

Almost the same as MLE, but with a prior distribution on the parameters

Estimating parameters of probability models

- In the occupancy grid mapping problem we wanted to compute $p(\mathbf{m}|\mathbf{z}_{1:t}, \mathbf{x}_{1:t})$ over all possible maps.

- We can see this problem as a specific instance within a category of problems where we are given data (observations) and we want to “explain” or fit the data using a parametric function.

- There are typically three ways to work with this type of problems:
    1. Maximum Likelihood parameter estimation (MLE)
        - Least Squares
    2. Maximum A Posteriori (MAP) parameter estimation
    3. Bayesian parameter distribution estimation

Bayesian parameter estimation

- Both MLE and MAP estimators give you a single **point estimate**.
- But there might be many parameters that are compatible with the data.
- Instead of point estimates, compute a **distribution of estimates** that explain the data

- Bayesian parameter estimation:

$$p(\boldsymbol{\theta}|\mathbf{d}_{1:N}) = \frac{p(\mathbf{d}_{1:N}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathbf{d}_{1:N})}$$

Note, this time we do not write “max”. So, dropping denominator is not trivial.

The probability of the data is usually hard to compute. But it does not depend on the parameter theta, so it is treated as a normalizing factor, and we can still compute how the posterior varies with theta.

Bayesian parameter estimation

- Both MLE and MAP estimators give you a single **point estimate**.
- But there might be many parameters that are compatible with the data.
- Instead of point estimates, compute a **distribution of estimates** that explain the data

- Bayesian parameter estimation:

$$p(\boldsymbol{\theta}|\mathbf{d}_{1:N}) = \frac{p(\mathbf{d}_{1:N}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathbf{d}_{1:N})}$$

- This is what we used in occupancy grid mapping, when we approximated

$$p(\mathbf{m}|\mathbf{z}_{1:t}, \mathbf{x}_{1:t})$$

Concluding thoughts: When to use each type of estimation?

- The computational complexity increases as we go through MLE, MAP and Bayesian

- Our reliability of solution also increases, as long as we have good priors!

- Major take-home: maximizing Gaussian likelihood equivalent to least squares solution!