22. Linear least squares problems

Last time

- Orthogonality
- Gram–Schmidt algorithm
- QR factorization

Goals for today

- Formulating data fitting problems
- Optimization
- Linear least squares problems
- Solution using linear algebra

Parametric function

- Given data points (t_i, y_i)
- Suppose have presumed (statistical) **model** for data

Parametric function

- Given data points (t_i, y_i)
- Suppose have presumed (statistical) **model** for data
- I.e. suspect related as $y = f_{\mathbf{p}}(x)$, with noise:

$$y_i = f_{\mathbf{p}}(t_i) + \epsilon_i$$

where ϵ_i represents the noise on observation i

Parametric function

- Given data points (t_i, y_i)
- Suppose have presumed (statistical) **model** for data
- I.e. suspect related as $y = f_{\mathbf{p}}(x)$, with noise:

$$y_i = f_{\mathbf{p}}(t_i) + \epsilon_i$$

  where ϵ_i represents the noise on observation i
- $f_{\mathbf{p}}$ is **parametric function** with parameter vector **p**

Parametric function

- Given data points (t_i, y_i)
- Suppose have presumed (statistical) **model** for data
- I.e. suspect related as $y = f_{\mathbf{p}}(x)$, with noise:

$$y_i = f_{\mathbf{p}}(t_i) + \epsilon_i$$

where ϵ_i represents the noise on observation i

- $f_{\mathbf{p}}$ is **parametric function** with parameter vector **p**
- Most common example: straight line $f_{\alpha,\beta}(x) = \alpha + \beta x$

Best fit

- Want to find **best fit** to the data
- How can we formalize this?

Best fit

- Want to find **best fit** to the data
- How can we formalize this?

- Find values of α, β that **minimize** distance of data from function
- **Parameter estimation** in statistics

Loss function

- Distance: **loss function** or **cost function** (optimization)
- E.g.

Loss function

- Distance: **loss function** or **cost function** (optimization)
- E.g.

$$\mathcal{L}(\alpha, \beta) := \sum_i \ell(y_i, f_{\alpha,\beta}(x_i))$$

where ℓ is a measure of distance

Loss function

- Distance: **loss function** or **cost function** (optimization)
- E.g.

$$\mathcal{L}(\alpha, \beta) := \sum_i \ell(y_i, f_{\alpha,\beta}(x_i))$$

where ℓ is a measure of distance

- Most common choice

$$\ell(x, y) = \|x - y\|_2^2$$

- **Least squares**

Least squares

- Want to minimize

$$\mathcal{L}(\mathbf{p}) = \sum_i \left[y_i - f_{\mathbf{p}}(t_i)\right]^2$$

Least squares

- Want to minimize

$$\mathcal{L}(\mathbf{p}) = \sum_i \left[y_i - f_{\mathbf{p}}(t_i) \right]^2$$

- Of form $\sum r_i^2$, where $r_i := y_i - f_{\mathbf{p}}(t_i)$ is ith **residual**
- Sum of squares, so minimum is when all $r_i = 0$

Least squares

- Want to minimize

$$\mathcal{L}(\mathbf{p}) = \sum_i \left[y_i - f_{\mathbf{p}}(t_i) \right]^2$$

- Of form $\sum r_i^2$, where $r_i := y_i - f_{\mathbf{p}}(t_i)$ is ith **residual**
- Sum of squares, so minimum is when all $r_i = 0$
- But data do not lie exactly on any $y = f_{\alpha,\beta}(x)$
- Choose curve that **minimizes** square sum of residuals

Least squares

- Want to minimize

$$\mathcal{L}(\mathbf{p}) = \sum_i \left[y_i - f_{\mathbf{p}}(t_i)\right]^2$$

- Of form $\sum r_i^2$, where $r_i := y_i - f_{\mathbf{p}}(t_i)$ is ith **residual**
- Sum of squares, so minimum is when all $r_i = 0$
- But data do not lie exactly on any $y = f_{\alpha,\beta}(x)$
- Choose curve that **minimizes** square sum of residuals
- **least squares** minimization
- Statistics: **linear regression**

Optimization

- Example of an **optimization** problem
- Optimization problems are usually **very hard**; huge field

Optimization

- Example of an **optimization** problem
- Optimization problems are usually **very hard**; huge field
- Special methods for problems with certain **structure**

Optimization

- Example of an **optimization** problem
- Optimization problems are usually **very hard**; huge field
- Special methods for problems with certain **structure**

- There are many numerical methods for optimization
- Sometimes analytical solutions are possible – e.g. linear least squares

Matrix formulation of linear least squares

- Look for a matrix formulation: $\mathcal{L} = \sum_i r_i^2 = r^T r$

Matrix formulation of linear least squares

- Look for a matrix formulation: $\mathcal{L} = \sum_i r_i^2 = r^T r$

$$\mathbf{r} = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_m \end{bmatrix} = \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \\ 1 & t_m \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} - \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} =: \mathsf{A}\mathbf{x} - \mathbf{b}$$

where $\mathbf{x} = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$ are the unknowns

Matrix formulation II

- Ideally would like to solve $Ax = b$

Matrix formulation II

- Ideally would like to solve $Ax = b$
- But have **overdetermined system**:
- Too many equations (n) for number of unknowns (2)

Matrix formulation II

- Ideally would like to solve $Ax = b$
- But have **overdetermined system**:
- Too many equations (n) for number of unknowns (2)

- Best we can do is minimize $r^T r$

General linear least squares

- General case: want to "solve" A$\mathbf{x} = \mathbf{b}$

  for **tall and narrow** matrix A

General linear least squares

- General case: want to "solve" A$\mathbf{x} = \mathbf{b}$

  for **tall and narrow** matrix A
- i.e. $(m \times n)$ matrix with $m > n$
- Again overdetermined system

General linear least squares

- General case: want to "solve" A**x** $=$ **b**

  for **tall and narrow** matrix A
- i.e. $(m \times n)$ matrix with $m > n$
- Again overdetermined system

- Best we can do is minimize

$$\|A\mathbf{x} - \mathbf{b}\|^2$$

General linear least squares

- General case: want to "solve" A**x** $=$ **b**

  for **tall and narrow** matrix A
- i.e. $(m \times n)$ matrix with $m > n$
- Again overdetermined system

- Best we can do is minimize

$$\|A\mathbf{x} - \mathbf{b}\|^2$$

- E.g. fit a polynomial of degree $< n$ to $(n+1)$ points

Geometrical interpretation

- Suppose $\mathbf{x}$ ranges over all of $\mathbb{R}^n$
- Then $\mathsf{A}\mathbf{x}$ ranges over column space / range of A

Geometrical interpretation

- Suppose **x** ranges over all of $\mathbb{R}^n$
- Then A**x** ranges over column space / range of A
- **Column space** of A: vec. space spanned by columns of A

Geometrical interpretation

- Suppose **x** ranges over all of $\mathbb{R}^n$
- Then A**x** ranges over column space / range of A
- **Column space** of A: vec. space spanned by columns of A

- Column space is a hyperplane

Geometrical interpretation

- Suppose **x** ranges over all of $\mathbb{R}^n$
- Then A**x** ranges over column space / range of A
- **Column space** of A: vec. space spanned by columns of A

- Column space is a hyperplane
- Looking for point **x** whose image A**x** is **closest** to b

Geometrical interpretation

- Suppose **x** ranges over all of $\mathbb{R}^n$
- Then A**x** ranges over column space / range of A
- **Column space** of A: vec. space spanned by columns of A

- Column space is a hyperplane
- Looking for point **x** whose image A**x** is **closest** to b
- Intuitively: when $r :=$ A**x** $-$ **b perpendicular** to column space

Solving linear least squares

- Want to find $\mathbf{x}$ minimizing $\|A\mathbf{x} - \mathbf{b}\|$

Solving linear least squares

- Want to find $\mathbf{x}$ minimizing $\|A\mathbf{x} - \mathbf{b}\|$
- Recall: $\|\mathbf{x}\|^2 = \mathbf{x}^\top \mathbf{x}$

Solving linear least squares

- Want to find $\mathbf{x}$ minimizing $\|A\mathbf{x} - \mathbf{b}\|$
- Recall: $\|\mathbf{x}\|^2 = \mathbf{x}^\top \mathbf{x}$
- Take any vector $\mathbf{y}$ and look at $\|A(\mathbf{x} + \mathbf{y}) - \mathbf{b}\|$:

Solving linear least squares

- Want to find $\mathbf{x}$ minimizing $\|A\mathbf{x} - \mathbf{b}\|$
- Recall: $\|\mathbf{x}\|^2 = \mathbf{x}^\top \mathbf{x}$
- Take any vector $\mathbf{y}$ and look at $\|A(\mathbf{x} + \mathbf{y}) - \mathbf{b}\|$:

$$\|(A\mathbf{x} - \mathbf{b}) + A\mathbf{y}\|^2 = [(A\mathbf{x} - \mathbf{b}) + A\mathbf{y}]^\top [(A\mathbf{x} - \mathbf{b}) + A\mathbf{y}]$$

Solving linear least squares

- Want to find $\mathbf{x}$ minimizing $\|A\mathbf{x} - \mathbf{b}\|$
- Recall: $\|\mathbf{x}\|^2 = \mathbf{x}^\top \mathbf{x}$
- Take any vector $\mathbf{y}$ and look at $\|A(\mathbf{x} + \mathbf{y}) - \mathbf{b}\|$:

$$\|(A\mathbf{x} - \mathbf{b}) + A\mathbf{y}\|^2 = [(A\mathbf{x} - \mathbf{b}) + A\mathbf{y}]^\top [(A\mathbf{x} - \mathbf{b}) + A\mathbf{y}]$$

$$= (A\mathbf{x} - \mathbf{b})^\top (A\mathbf{x} - \mathbf{b}) + 2\mathbf{y}^\top A^\top (A\mathbf{x} - \mathbf{b}) + (A\mathbf{y})^\top (A\mathbf{y})$$

Solving linear least squares

- Want to find $\mathbf{x}$ minimizing $\|A\mathbf{x}-\mathbf{b}\|$
- Recall: $\|\mathbf{x}\|^2 = \mathbf{x}^\top \mathbf{x}$
- Take any vector $\mathbf{y}$ and look at $\|A(\mathbf{x}+\mathbf{y})-\mathbf{b}\|$:

$$\|(A\mathbf{x}-\mathbf{b})+A\mathbf{y}\|^2 = [(A\mathbf{x}-\mathbf{b})+A\mathbf{y}]^\top[(A\mathbf{x}-\mathbf{b})+A\mathbf{y}]$$

$$= (A\mathbf{x}-\mathbf{b})^\top(A\mathbf{x}-\mathbf{b}) + 2\mathbf{y}^\top A^\top(A\mathbf{x}-\mathbf{b}) + (A\mathbf{y})^\top(A\mathbf{y})$$

- Here used $\mathbf{y}^\top\mathbf{z} = \mathbf{z}^\top\mathbf{y}$, so $\mathbf{y}^\top\mathbf{z}+\mathbf{z}^\top\mathbf{y} = 2\mathbf{y}^\top\mathbf{z}$

$$= \|A\mathbf{x}-\mathbf{b}\|^2 + \|A\mathbf{y}\|^2 + 2\mathbf{y}^\top A^\top(A\mathbf{x}-\mathbf{b})$$

Solving linear least squares II

- Arrived at $\|A\mathbf{x} - \mathbf{b}\|^2 + \|A\mathbf{y}\|^2 + 2\mathbf{y}^\top A^\top (A\mathbf{x} - \mathbf{b})$

Solving linear least squares II

- Arrived at $\|A\mathbf{x} - \mathbf{b}\|^2 + \|A\mathbf{y}\|^2 + 2\mathbf{y}^\top A^\top (A\mathbf{x} - \mathbf{b})$
- Take $\mathbf{x}$ such that $A^\top (A\mathbf{x} - \mathbf{b}) = 0$

Solving linear least squares II

- Arrived at $\|A\mathbf{x} - \mathbf{b}\|^2 + \|A\mathbf{y}\|^2 + 2\mathbf{y}^\top A^\top (A\mathbf{x} - \mathbf{b})$
- Take $\mathbf{x}$ such that $A^\top (A\mathbf{x} - \mathbf{b}) = 0$
- Then for any $\mathbf{y}$ we have

$$\|A(\mathbf{x} + \mathbf{y}) - \mathbf{b}\|^2 \geq \|A\mathbf{x} - \mathbf{b}\|^2$$

Solving linear least squares II

- Arrived at $\|A\mathbf{x} - \mathbf{b}\|^2 + \|A\mathbf{y}\|^2 + 2\mathbf{y}^\top A^\top (A\mathbf{x} - \mathbf{b})$
- Take $\mathbf{x}$ such that $A^\top (A\mathbf{x} - \mathbf{b}) = 0$
- Then for any $\mathbf{y}$ we have

$$\|A(\mathbf{x} + \mathbf{y}) - \mathbf{b}\|^2 \geq \|A\mathbf{x} - \mathbf{b}\|^2$$

- So $\mathbf{x}$ is minimizer

Solving linear least squares II

- Arrived at $\|A\mathbf{x} - \mathbf{b}\|^2 + \|A\mathbf{y}\|^2 + 2\mathbf{y}^\top A^\top (A\mathbf{x} - \mathbf{b})$
- Take $\mathbf{x}$ such that $A^\top (A\mathbf{x} - \mathbf{b}) = 0$
- Then for any $\mathbf{y}$ we have

$$\|A(\mathbf{x} + \mathbf{y}) - \mathbf{b}\|^2 \geq \|A\mathbf{x} - \mathbf{b}\|^2$$

- So $\mathbf{x}$ is minimizer
- Unique solution of least squares problem given by

$$A^\top A\mathbf{x} = A^\top \mathbf{b}$$

Solving linear least squares III

- Reduced linear least squares problem to *linear* equation:

$$\mathsf{A}^\top \mathsf{A} x = \mathsf{A}^\top \mathbf{b}$$

  – called **normal equations**

- Square matrix $\mathsf{A}^\top \mathsf{A}$

Solving linear least squares III

- Reduced linear least squares problem to *linear* equation:

$$\mathsf{A}^\top \mathsf{A} x = \mathsf{A}^\top \mathbf{b}$$

– called **normal equations**

- Square matrix $\mathsf{A}^\top \mathsf{A}$
- Solvable if A is **full rank**, i.e. columns span the space

Pseudoinverse

- Mathematically get $\mathbf{x} = (\mathsf{A}^\top \mathsf{A})^{-1} \mathsf{A}^\top \mathbf{b}$

Pseudoinverse

- Mathematically get $\mathbf{x} = (\mathsf{A}^\top \mathsf{A})^{-1} \mathsf{A}^\top \mathbf{b}$
- Notation: $\mathbf{x} = \mathsf{A}^+ \mathbf{b}$

- Where $\mathsf{A}^+ := (\mathsf{A}^\top \mathsf{A})^{-1} \mathsf{A}^\top$ is **pseudo-inverse**

Pseudoinverse

- Mathematically get $\mathbf{x} = (\mathsf{A}^\top \mathsf{A})^{-1} \mathsf{A}^\top \mathbf{b}$
- Notation: $\mathbf{x} = \mathsf{A}^+ \mathbf{b}$

- Where $\mathsf{A}^+ := (\mathsf{A}^\top \mathsf{A})^{-1} \mathsf{A}^\top$ is **pseudo-inverse**
- $\mathsf{A}^\top \mathsf{A}$ is invertible iff A has full rank

Solving normal equations

- Possible to solve normal equations using elimination, but to do so need to construct $A^\top A$

Solving normal equations

- Possible to solve normal equations using elimination, but to do so need to construct $A^\top A$
- Slightly more efficient version of elimination for **symmetric** matrices: **Cholesky factorization**

Solving normal equations

- Possible to solve normal equations using elimination, but to do so need to construct $A^\top A$
- Slightly more efficient version of elimination for **symmetric** matrices: **Cholesky factorization**
- Alternative: factorize $A = QR$

Solving normal equations

- Possible to solve normal equations using elimination, but to do so need to construct $\mathsf{A}^\top \mathsf{A}$
- Slightly more efficient version of elimination for **symmetric** matrices: **Cholesky factorization**

- Alternative: factorize $\mathsf{A} = \mathsf{Q}\mathsf{R}$
- Need generalization of QR when A tall and narrow, $m \times n$

Solving normal equations

- Possible to solve normal equations using elimination, but to do so need to construct $\mathsf{A}^\top \mathsf{A}$
- Slightly more efficient version of elimination for **symmetric** matrices: **Cholesky factorization**

- Alternative: factorize $\mathsf{A} = \mathsf{QR}$
- Need generalization of QR when A tall and narrow, $m \times n$
- **Thin** QR factorization: $A = QR$ where
  - Q is $(m \times n)$ with orthonormal columns
  - R is $(n \times n)$ upper triangular

Solving normal equations

- Possible to solve normal equations using elimination, but to do so need to construct $\mathsf{A}^\top \mathsf{A}$
- Slightly more efficient version of elimination for **symmetric** matrices: **Cholesky factorization**

- Alternative: factorize $\mathsf{A} = \mathsf{QR}$
- Need generalization of QR when A tall and narrow, $m \times n$
- **Thin** QR factorization: $A = QR$ where
  - Q is $(m \times n)$ with orthonormal columns
  - R is $(n \times n)$ upper triangular
- If A is of full rank then R is non-singular

Solving normal equations II

- Want to solve $A^\top A = A^\top \mathbf{b}$
- Substitute $A = QR$

Solving normal equations II

- Want to solve $A^\top A = A^\top \mathbf{b}$
- Substitute $A = QR$

- So $R^\top Q^\top QR\,\mathbf{x} = R^\top Q^\top \mathbf{b}$

Solving normal equations II

- Want to solve $A^\top A = A^\top \mathbf{b}$
- Substitute $A = QR$

- So $R^\top Q^\top QR\,\mathbf{x} = R^\top Q^\top \mathbf{b}$
- $R^\top R\,\mathbf{x} = R^\top Q^\top \mathbf{b}$

Solving normal equations II

- Want to solve $A^\top A = A^\top \mathbf{b}$
- Substitute $A = QR$

- So $R^\top Q^\top QR\,\mathbf{x} = R^\top Q^\top \mathbf{b}$
- $R^\top R\,\mathbf{x} = R^\top Q^\top \mathbf{b}$
- $R\,\mathbf{x} = Q^\top \mathbf{b}$

Solving normal equations II

- Want to solve $\mathsf{A}^\top\mathsf{A} = \mathsf{A}^\top\mathbf{b}$
- Substitute $\mathsf{A} = \mathsf{QR}$

- So $\mathsf{R}^\top\mathsf{Q}^\top\mathsf{QR}\,\mathbf{x} = \mathsf{R}^\top\mathsf{Q}^\top\mathbf{b}$
- $\mathsf{R}^\top\mathsf{R}\,\mathbf{x} = \mathsf{R}^\top\mathsf{Q}^\top\mathbf{b}$
- $\mathsf{R}\,\mathbf{x} = \mathsf{Q}^\top\mathbf{b}$
- So $\mathbf{x} = \mathsf{R}^{-1}\mathsf{Q}^\top b$

Solving normal equations II

- Want to solve $A^\top A = A^\top \mathbf{b}$
- Substitute $A = QR$

- So $R^\top Q^\top QR\,\mathbf{x} = R^\top Q^\top \mathbf{b}$
- $R^\top R\,\mathbf{x} = R^\top Q^\top \mathbf{b}$
- $R\,\mathbf{x} = Q^\top \mathbf{b}$
- So $\mathbf{x} = R^{-1} Q^\top b$

- Solve $R\mathbf{x} = Q^\top \mathbf{b}$ by backsubstitution

Backslash

- Backslash in Julia is overloaded to give this least-squares solution when A is an $(m \times n)$ matrix
- It is not necessarily clear that this is a good idea!

Backslash

- Backslash in Julia is overloaded to give this least-squares solution when A is an $(m \times n)$ matrix
- It is not necessarily clear that this is a good idea!

- E.g. simple linear fit: use \ with above matrix

Summary

- Linear least squares for overdetermined systems
- Solvable using linear algebra
- Solution given by normal equations – linear system
- Solve using QR decomposition