I. Function approximation with given points.

Given points. \( \{ (x_i, y_i) \}_{i=0}^{n} \) (which are assumed to be samples of a function \( f \)). Find a function \( p \) such that

\[
p(x_i) = y_i \quad i = 0, 1, \ldots, n.
\]

II. Regression in statistical learning theory.

\( x \) is sampled from a probability measure in \( \mathbb{R}^d \).

\[
y = f(x) + \xi.
\]

Where \( \xi \) is a random noise independent of \( x \).

Given samples. \( \{ (x_i, y_i) \}_{i=1}^{n} \), construct a function \( \hat{f} \).

Error \( \| \hat{f} - f \|_2 \).

Mean Square Error: \( \mathbb{E} \| \hat{f} - f \|_2^2 \rightarrow 0 \) as \( n \rightarrow \infty \).

Rate of convergence.

Interpolation

Given \( n+1 \) data points. \( \{ (x_i, y_i) \}_{i=1}^{n} \), we want to express an interpolation function \( f(x) \) as a linear combination of a set of basis functions \( \{ q_0, q_1, \ldots, q_n \} \) so that

\[
f(x) \approx c_0 q_0(x) + c_1 q_1(x) + \cdots + c_n q_n(x)
\]

where \( c_0, c_1, \ldots, c_n \) are to be determined.

Need. \( f(x_i) = c_0 q_0(x_i) + c_1 q_1(x_i) + \cdots + c_n q_n(x_i) = y_i \) for \( i = 0, 1, \ldots, n \).
Polynomial interpolation.

Natural basis. \( q_0(x) = 1 \) \( q_1(x) = x \) \( q_2(x) = x^2 \) \( \ldots \) \( q_n(x) = x^n \).

Polynomial of degree \( \leq n \). \( P_n(x) = c_0 + c_1 x + c_2 x^2 + \ldots + c_n x^n \).

Set. \( P_n(x_i) = y_i \quad i = 0, 1, \ldots, n \).

\[
\begin{bmatrix}
1 & x_0 & x_0^2 & \ldots & x_0^n \\
1 & x_1 & x_1^2 & \ldots & x_1^n \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
1 & x_n & x_n^2 & \ldots & x_n^n \\
\end{bmatrix}
\begin{bmatrix}
c_0 \\
c_1 \\
\vdots \\
c_n \\
\end{bmatrix}
=
\begin{bmatrix}
y_0 \\
y_1 \\
\vdots \\
y_n \\
\end{bmatrix}
\]

Vandermonde matrix denoted by \( V_n \).

Thm 1 (Determinant of \( V_n \)). For a non-negative integer \( n \), and any sequence \( \{x_0, \ldots, x_n\} \) of \( n+1 \) points in \( \mathbb{R} \),

\[
\det(V_n) = \prod_{j=0}^{n-1} \prod_{k=j+1}^{n} (x_k - x_j).
\]

- \( V_n \) is invertible if \( \{x_0, x_1, \ldots, x_n\} \) are distinct.
- There exists a unique \( \mathbf{c} = \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_n \end{bmatrix} \) such that \( V_n \mathbf{c} = \mathbf{y} \).
- There exists a unique polynomial with degree \( \leq n \) that interpolates \( \{(x_i, y_i) : y_i = i \text{ if} \{x_0, x_1, \ldots, x_n\} \text{ are distinct.} \}

Thm 2. If \( \{x_0, x_1, \ldots, x_n\} \) are distinct real numbers, then for arbitrary values \( y_0, y_1, \ldots, y_n \), there is a unique polynomial \( P_n \) of degree at most \( n \) such that

\[ P_n(x_i) = y_i \quad i = 0, 1, \ldots, n. \]
Another proof of Theorem 2, resulting in Newton's Interpolation Formula.

1. Interpolate \((x_0, y_0)\), \(\deg 0 \): \(P_0(x) = c_0\) \(c_0 = y_0\).

2. Interpolate \((x_0, y_0)\), \(\deg 1 \): \(P_1(x) = P_0(x) + c_1(x-x_0)\) \(\text{Need } P_1(x) = P_0(x) + c_1(x-x_0)\) \(\text{to choose } c_1 :\)

3. Interpolate \((x_0, y_0)\), \(\deg 2 \): \(P_2(x) = P_1(x) + c_2(x-x_0)(x-x_1)\). \(\text{Choose } c_2 \text{ such that } P_2(x_2) = P_1(x_2) + c_2(x_2-x_0)(x_2-x_1)\).

Obtain \(P_{k-1}(x)\), \(\deg \leq k-1\) that interpolates \((x_0, y_0) \ldots (x_{k-1}, y_{k-1})\).

Construct \(P_k(x) = P_{k-1}(x) + c_k(x-x_0) \ldots (x-x_{k-1})\), \(\text{of } \deg \leq k\).

Choose \(c_k\) such that \(P_k(x_k) = P_{k-1}(x_k) + c_k(x_k-x_0) \ldots (x_k-x_{k-1})\).

Then \(P_k(x)\) interpolates \((x_0, y_0) \ldots (x_k, y_k)\).

Finally, we obtain \(P_n(x)\), \(\deg \leq n\) that interpolates \(\{(x_i, y_i)\}_{i=0}^n\).

\[P_0(x) = c_0,\]
\[P_1(x) = P_0(x) + c_1(x-x_0)\]
\[P_2(x) = P_1(x) + c_2(x-x_0)(x-x_1),\]
\[P_k(x) = P_{k-1}(x) + c_k(x-x_0) \ldots (x-x_{k-1}),\]
\[P_n(x) = \sum_{i=0}^{k} c_i \prod_{j=0}^{i-1} (x-x_j), \text{ where } \prod_{j=0}^{m} (x-x_j) = 1 \text{ whenever } m < 0.\]

**Newton form of Interpolation Polynomial**
How to evaluate $P_k(x)$ efficiently if the coefficients $c_0, c_1, \ldots, c_k$ are known?

$$P_k(x) = \sum_{i=0}^{k} c_i \prod_{j=0}^{i-1} (x - x_j)$$

Let $d_0 = x - x_0$, $d_1 = x - x_1$, $\ldots$

$$= \sum_{i=0}^{k} c_i \prod_{j=0}^{i-1} d_j = \sum_{i=0}^{k} c_i d_0 d_1 \ldots d_{i-1}.$$  

$$= c_0 + c_1 d_0 + c_2 d_0 d_1 + c_3 d_0 d_1 d_2 + \ldots + c_{k-1} d_0 d_1 \ldots d_{k-2} + c_k d_0 d_1 \ldots d_{k-1}.$$  

$$= \left( \left( c_k d_0 + c_{k-1} \right) d_2 + c_{k-2} \right) d_3 + \ldots + c_1 \right) d_0 + c_0.$$  

Algorithm:

$$U_k \leftarrow c_k,$$

$$U_{k-1} \leftarrow U_k d_{k-1} + c_{k-1}.$$

$$U_{k-2} \leftarrow U_{k-1} d_{k-2} + c_{k-2}.$$

$$\vdots$$

$$U_0 \leftarrow U_1 d_0 + c_0.$$  

How to compute the coefficients $c_i$.

$$c_k = \frac{y_k - P_{k-1}(x_k)}{(x_k - x_0)(x_k - x_1) \ldots (x_k - x_{k-1})}$$
Lagrange Form of the Interpolation Polynomial

\[ p(x) = y_0 \cdot l_0(x) + y_1 \cdot l_1(x) + \cdots + y_n \cdot l_n(x) = \sum_{k=0}^{n} y_k \cdot l_k(x) \]

\( l_0(x), l_1(x), \ldots, l_n(x) \) are polynomials depending on the nodes \( x_0, x_1, \ldots, x_n \).

\[ l_i(x) = \prod_{j=0 \atop j \neq i}^{n} \frac{x - x_i}{x_i - x_j}, \quad 0 \leq i \leq n. \]

For example, \( l_0(x) = \frac{\prod_{j=1}^{n} \frac{x - x_j}{x_0 - x_j}}{x_0 - x_j} \).

These polynomials satisfy \( l_i(x_j) = \delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases} \)

Then \( p(x) = \sum_{k=0}^{n} y_k \cdot l_k(x) = y_i \), so \( p(x) \) interpolates \( \{(x_i, y_i)\}_{i=0}^{n} \).

Example:

<table>
<thead>
<tr>
<th>( x )</th>
<th>5</th>
<th>-7</th>
<th>-6</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>( y )</td>
<td>1</td>
<td>-23</td>
<td>-54</td>
<td>-954</td>
</tr>
</tbody>
</table>

\( x_0 = 5 \quad x_1 = -7 \quad x_2 = -6 \quad x_3 = 0 \).

\( l_0(x) = \frac{(x+7)(x+6)x}{(5+7)(5+6)5} = \frac{1}{660} x(x+6)(x+7) \).  
\( l_1(x) = \frac{(x-5)(x+6)x}{(-7-5)(-7+6)(-7)} = \frac{-1}{840} x(x-5)(x+6) \).  
\( l_2(x) = \frac{(x-5)(x+7)x}{(-6-5)(-6+7)(-6)} = \frac{-1}{660} x(x-5)(x+7) \).  
\( l_3(x) = \frac{(x-5)(x+7)(x+6)}{(-5)(7)(6)} = \frac{-1}{210} x(x-5)(x+6)(x+7) \).  

Interpolating polynomial \( p(x) = l_0(x) - 23l_1(x) - 54l_2(x) - 954l_3(x) \).
Three polynomial interpolation forms.

I. Monomial basis. Resulting a linear system of Vandermonde matrices.
   - The Vandermonde matrix is often ill-conditioned.
   - Not suitable for computation.

II. Newton form.
   - Best for computation.
   - If more data points are added to the interpolation problem, the coefficients already computed will not have to be changed.
   - Easily accommodates additional data to be interpolated.

III. Lagrangean form.

The Error in Polynomial Interpolation.

Let \( f \) be a function on \([a, b]\) such that \( \{x_i, y_i\}_{i=0}^{n} \) samples of the function such that \( y_i = f(x_i) \). \( P(x) \) the polynomial of degree \( \leq n \) that interpolates \( f \) at \( n+1 \) distinct points \( x_0, x_1, \ldots, x_n \) in the interval \([a, b] \). To each \( x \in [a, b] \), there corresponds a point \( s_x \in [a, b] \) such that

\[
f(x) - p(x) = \frac{1}{(n+1)!} \sum_{i=0}^{n} \frac{f^{(n+1)}}{n!} (s_x) (x - x_i)
\]

Theorem 3. [Theorem on Polynomial Interpolation Error]

\[
f(x) - p(x) = \frac{1}{(n+1)!} \sum_{i=0}^{n} \frac{f^{(n+1)}}{n!} (s_x) (x - x_i).
\]
Proof.

If $x = x_0$, or $x_1$ or ... $x_n$, Eq. (1) is obviously true.

Assume $x$ is any point other than a node. Let $x$ be fixed.

Define $W(t) = \prod_{i=0}^{n} (t - x_i), \quad \omega(x) = \prod_{i=0}^{n} (x - x_i)$

$\phi := f - P - \lambda x \omega$ \quad $\lambda x$ is set such that 

$$\lambda x = \frac{f(x) - P(x)}{\omega(x)}.$$

$\phi \in C^{n+1} [a,b]. \quad \phi(x) = 0, \quad \phi(x_0) = 0 \ldots \phi(x_n) = 0.$

Rolle's theorem. $\phi'$ has at least $n+1$ distinct zeros in $(a,b)$.

$\phi'' \quad \ldots \quad \llap{n} \quad \ldots \quad \llap{1} \quad \text{zero is } (a,b)$

Let it be $s_x$.

$\phi^{(n+1)} (s_x) = f^{(n+1)} (s_x) - P^{(n+1)} (s_x) - \lambda x \omega^{(n+1)} (s_x)$

$\lambda x = \frac{f^{(n+1)} (s_x)}{(n+1)!}$

Example. $f(x) = \sin x$. Is approximated by a polynomial of degree $9$
that interpolates $f$ at 10 points in the interval $[0,1]$. How large is the error?

$|\sin x - P(x)| \leq \frac{1}{10!} \left| f^{(10)} (s_x) \prod_{i=0}^{9} (s_x - x_i) \right| \leq \frac{1}{10!}$

$P^{(10)}(x) \leq 1 \left| \prod_{i=0}^{9} (x - x_i) \right| \leq 1$
Chebyshev Polynomials

Chebyshev polynomials are defined recursively.

\[ T_0(x) = 1 \quad T_1(x) = x \quad T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x) \]

Such that:

\[ T_2(x) = 2x^2 - 1 \]
\[ T_3(x) = 4x^3 - 3x \]
\[ T_4(x) = 8x^4 - 8x^2 + 1 \]
\[ T_5(x) = 16x^5 - 20x^3 + 5x \]

Closed form of Chebyshev polynomials for \( x \in [-1, 1] \):

Theorem 4. For \( x \in [-1, 1] \), the Chebyshev polynomials have the closed form expression.

\[ T_n(x) = \cos(n \cos^{-1} x) \quad n \geq 0 \]

Proof: Recall \( \cos(A + B) = \cos A \cos B - \sin A \sin B \).

Then:

\[ \cos(n+1)\theta = \cos\theta \cos n\theta - \sin\theta \sin n\theta \]
\[ \cos(n-1)\theta = \cos\theta \cos n\theta + \sin\theta \sin n\theta \]

Adding these two equations gives:

\[ \cos(n+1)\theta + \cos(n-1)\theta = 2 \cos n\theta \cos\theta \]

\[ \Rightarrow \cos(n+1)\theta = 2 \cos n\theta \cos\theta - \cos(n-1)\theta \]

Let \( x = \cos \theta \). \( f_n(x) = \cos n\theta \) \( = \cos(n \cos^{-1} x) \)

\( f_0(x) = 1 \)
\( f_1(x) = x \)
\( f_{n+1}(x) = 2x f_n(x) - f_{n-1}(x) \)

So \( f_n(x) = T_n(x) \)
Properties of Chebyshev polynomials.

I. \[ |T_n(x)| \leq 1 \quad x \in [-1, 1] \]

II. \[ T_n \left( \cos \frac{j\pi}{n} \right) = (-1)^j \quad 0 \leq j \leq n. \]

\[ T_n \left( \cos \frac{2j-1}{2n} \pi \right) = 0 \quad 1 \leq j \leq n. \]

III. \[ T_n(x) = 2^{n-1} \prod_{j=1}^{n} \left[ x - \cos \frac{2j-1}{2n} \pi \right]. \]

The Chebyshev polynomial, \( T_n(x) \), is of degree \( n \). The highest order term is \( 2^{n-1}x^n \).

It has \( n \) distinct roots, \( \cos \frac{2j-1}{2n} \pi \).
Approximation error by infinity norm.
\[ \| f - P \|_\infty = \max_{x \in [a,b]} | f(x) - P(x) |. \]

According to Theorem 3, \[ | f(x) - P(x) | \leq \frac{1}{(n+1)!} \left( \frac{1}{n} \right) \left| f^{(n+1)}(\xi) \right| \max_{x \in [a,b]} \left| \frac{1}{i=0} \sum_{i=0}^{n} (x-x_i) \right|, \]
where \( \xi \in [a,b] \).

If \( x \in [-1,1] \),
\[ \| f - P \|_\infty \leq \frac{1}{(n+1)!} \left( \frac{1}{n} \right) \max_{x \in [-1,1]} \left| f^{(n+1)}(\xi) \right| \max_{x \in [-1,1]} \left| \frac{1}{i=0} \sum_{i=0}^{n} (x-x_i) \right|. \]

Theorem on Monic polynomials.
A monic polynomial is one in which the term of highest degree has a coefficient of unity.

Theorem 4. If \( p \) is a monic polynomial of degree \( n \), then
\[ \| P \|_\infty = \max_{x \in [-1,1]} | P(x) | \geq 2^{1-n}. \]

Error analysis.
Let \( Q_{n+1}(x) = \frac{1}{(n+1)!} \sum_{i=0}^{n} (x-x_i) \)
\[ \| Q_{n+1} \|_\infty = \max_{x \in [-1,1]} | Q_{n+1}(x) |. \]

Then \( \| Q_{n+1} \|_\infty \geq 2^{1-(n+1)} = 2^{-n} \) by Theorem 4.

What nodes \( \{x_0, x_1, \ldots, x_n\} \) give rise to the minimum \( \| Q_{n+1} \|_\infty \)?

- \( 2^{-n} T_{n+1} \) is a monic polynomial of degree \( n+1 \).
- \( \max_{x \in [-1,1]} | T_{n+1}(x) | = 1 \).
\[ \max_{x \in [-1,1]} \left| 2^{-n} T_{n+1}(x) \right| = 2^{-n}. \]

\[ \min_{x_0, x_1, \ldots, x_n \in [-1,1]} \| Q_{n+1} \|_{\infty} = \min_{x \in [-1,1]} \left| 2^{-n} T_{n+1}(x) \right| = \min_{x \in [-1,1]} \left| \sum_{i=0}^{n} (x - x_i^c) \right| \]

where \( x_i^c = \cos \left( \frac{2i+1}{2n+2} \pi \right) \quad i = 0, \ldots, n \) are the roots of \( T_{n+1} \).

In other words, \( \| Q_{n+1} \|_{\infty} \) is minimized if the nodes \( x_0, \ldots, x_n \) are taken to be the roots of \( T_{n+1} \).

\[ \text{Chebyshev nodes} : \text{ the roots of the Chebyshev polynomials} \]

**Theorem 5.** If the nodes \( x_0, x_1, \ldots, x_n \) are the roots of the Chebyshev polynomial \( T_{n+1} \). Let \( f \) be a function in \( C^{n+1} [-1,1] \) and let \( p \) be the polynomial of degree \( n \) that interpolates \( f \) at \( x_0, \ldots, x_n \). Then,

\[ \| f - p \|_{\infty} = \sup_{x \in [-1,1]} |f(x) - p(x)| \leq \frac{1}{2^{n+1}} \max_{x \in [-1,1]} |f^{(n+1)}(x)|. \]

**Remark.** Chebyshev polynomials and Chebyshev nodes are also defined on \([a, b]\) as well. Use the map:

\[ x = \frac{1}{2} (b-a) t + \frac{1}{2} (a+b) \quad t \in [-1,1] \]

or equivalently,

\[ t = \frac{2}{b-a} \left[ x - \frac{1}{2} (a+b) \right] \quad x \in [a, b] \]
If \( f \in C^{n+1}([a,b]) \), and \( x_0, x_1, \ldots, x_n \) are \( n+1 \) Chebyshev nodes on \([a,b]\). Let \( p \) be the polynomial of degree \( n \) that interpolates \( f \) at \( x_0, x_1, \ldots, x_n \). Then,

\[
\|f - p\|_{\infty} = \sup_{x \in [a,b]} |f(x) - p(x)| \leq \frac{1}{2n(n+1)!} \left( \frac{b-a}{2} \right)^{n+1} \|f^{(n+1)}\|_{\infty}
\]

Uniform convergence

\( \frac{1}{n} \) points

\[ \|f - P_n\|_{\infty} = \max_{x \in [a,b]} |f(x) - P_n(x)| \to 0 \quad \text{as} \quad n \to \infty \]

- The result may depend on the interpolation nodes.
- If Chebyshev nodes are used, \( \|f - P_n\|_{\infty} \to 0 \) if \( \|f^{(n)}\|_{\infty} \leq M \) for any \( n \).

**Example:** Runge example.

\[
f(x) = \frac{1}{1 + 25x^2} \quad x \in [\pi, 0] \]

Nodes: \( x_i = \frac{2i}{n} - 1 \quad i = \{0, 1, \ldots, n\} \)

\( P_n(x) \): Interpolating polynomials of degree \( \leq n \).

\[
\|f - P_n\|_{\infty} = \max_{x \in [-1, 1]} |f(x) - P_n(x)| \to \infty \quad \text{as} \quad n \to \infty
\]

Why?

\[
\max_{x \in [-1, 1]} |f(x) - P_n(x)| \leq \max_{x \in [-1, 1]} \frac{|f^{(n+1)}(x)|}{(n+1)!} \max_{x \in [-1, 1]} |x - x_i|,
\]

\( Q_{n+1}(x) = (x - x_0) \cdots (x - x_n) \)

\[
\max_{x \in [-1, 1]} |Q_{n+1}(x)| \leq h \cdot h^2 \cdot \cdots \cdot nh = n! \cdot h^{n+1}
\]

\( h = \frac{2}{n} \)}
\[ \| f - P_n \|_{\infty} \leq \frac{n! \cdot h^{n+1}}{(n+1)!} \max_{x \in [a,b]} |f^{(n+1)}(x)| = \frac{M_n h^{n+1}}{n+1} \]

**Problem:** \( M_n \) increases when \( n \) increases, very fast.

However, if \( P_n \) are interpolated at Chebyshev nodes,
\[ \| f - P_n \|_{\infty} \to 0 \quad \text{as} \quad n \to \infty. \]

**Classical results.**

**Thm1.** Faber’s Theorem.

For any prescribed system of nodes,
\[ a \leq x^{(n)}_0 \leq x^{(n)}_1 \leq \ldots \leq x^{(n)}_n \leq b \quad n \geq 0, \]
there exists a continuous function \( f \) on \([a,b]\) such that the interpolating polynomials for \( f \) using these nodes fail to converge uniformly to \( f \).

**Thm2.** If \( f \) is a continuous function on \([a,b]\), then there is a system of nodes such that the interpolating polynomials for \( f \) at these nodes satisfy \( \lim_{n \to \infty} \| f - P_n \|_{\infty} = 0 \).

**Theorem 2.** is obtained by joining the Weier–Stress Approximation Thm
and the Chebyshev Alternation Thm.
Thm. Weierstrass Approximation Theorem.

If \( f \) is continuous on \([a, b]\) and if \( \varepsilon > 0 \), then there is a polynomial \( P \) satisfying \( |f(x) - P(x)| < \varepsilon \) on the interval \([a, b]\).

Proof. One can prove it on \([0, 1]\) without loss of generality. On \([0, 1]\), \( f \in C([0, 1]) \), construct a sequence of Bernstein polynomials \( B_{n}f \) that converges to \( f \) uniformly.

Bernstein Polynomials:

\[
(B_{n}f)(x) = \sum_{k=0}^{n} \binom{n}{k} x^{k}(1-x)^{n-k} f(\frac{k}{n})
\]

\( B_{n} \) is a linear operator. \( B_{n}(af+bg) = aB_{n}f + bB_{n}g \).

Detailed proof is in. Kincaid & Cheney.