Math/Stat 2300 Smoothing (4.3): Low-Order Polynomial Models
from text *A First Course in Mathematical Modeling*, Giordano, Fox, Horton, Weir, 2009.

We want to find methods that retain the advantages of the higher-order polynomials
without the disadvantages.
One technique is to choose a low-order polynomial regardless of the number of data points
(then the number of data points exceeds the number of constants). The polynomial will
not pass through all data points, so a choice with regards to best fit must be decided.
For example, suppose we had a set of 10 data points and we had decided to fit a quadratic.
It is impossible to force a quadratic to pass through 10 data points, it must be decided
which quadratic best fits the data.
This process is called **smoothing**. This smoothing process requires two decisions: the
order of the interpolating polynomial must be selected and the coefficients of the
polynomial must be determined according to some criterion for the best-fitting polynomial.
Resulting from this we get an optimization problem as we discussed before.

Divided Differences (also called Newton Divided Differences)

The idea of divided differences is to use data points to estimate the slopes of our
polynomial.
Some notation:

$f[x_i]$ is the **zeroth divided difference**. This is just the y value at x_i

$f[x_i, x_{i+1}]$ is the **first divided difference**

$$f[x_i, x_{i+1}] = \frac{f[x_{i+1}] - f[x_i]}{x_{i+1} - x_i}$$

$f[x_i, x_{i+1}, x_{i+2}]$ is the **second divided difference**

$$f[x_i, x_{i+1}, x_{i+2}] = \frac{f[x_{i+1}, x_{i+2}] - f[x_i, x_{i+1}]}{x_{i+2} - x_i}$$

Continuing on, the k**th divided difference** is

$$f[x_i, x_{i+1}, \ldots, x_{i+k}] = \frac{f[x_{i+1}, \ldots, x_{i+k}] - f[x_i, \ldots, x_{i+k-1}]}{x_{i+k} - x_i}$$

Using these, the nth degree interpolating polynomial for the data points (x_i, y_i) (that is,
$(x_i, f[x_i])$) is given as

$$P_n(x) = f[x_0] + f[x_0, x_1](x - x_0) + f[x_0, x_1, x_2](x - x_0)(x - x_1) + \ldots$$

$$+ f[x_0, x_1, \ldots, x_n](x - x_0)(x - x_1) \cdots (x - x_{n-1})$$

Using a difference table we can easily calculate these coefficients.

Example

Consider the data:

$$
\begin{array}{c|ccccc}
x_i & 0 & 2 & 4 & 6 & 8 \\
\hline
y_i & 0 & 4 & 16 & 36 & 64
\end{array}
$$

A difference table for this data is given as follows:

Data		Divided Differences		
x_i $\quad f[x_i]$	$f[x_i, x_{i+1}]$	$f[x_i, x_{i+1}, x_{i+2}]$	$f[x_i, x_{i+1}, x_{i+2}, x_{i+3}]$	
0 $\quad$ 0				
	$\frac{4-0}{2-0} = 2$			
2 $\quad$ 4		$\frac{6-2}{4-0} = 1$		
	$\frac{16-4}{4-2} = 6$		0	
4 $\quad$ 16		$\frac{10-6}{6-2} = 1$		
	$\frac{36-16}{6-4} = 10$		0	
6 $\quad$ 36		$\frac{14-10}{8-4} = 1$		
	$\frac{64-36}{8-6} = 14$			
8 $\quad$ 64				

Then our interpolating polynomial is

$$
\begin{aligned}
P_2(x) &= f[x_0] + f[x_0, x_1](x - x_0) + f[x_0, x_1, x_2](x - x_0)(x - x_1) \\
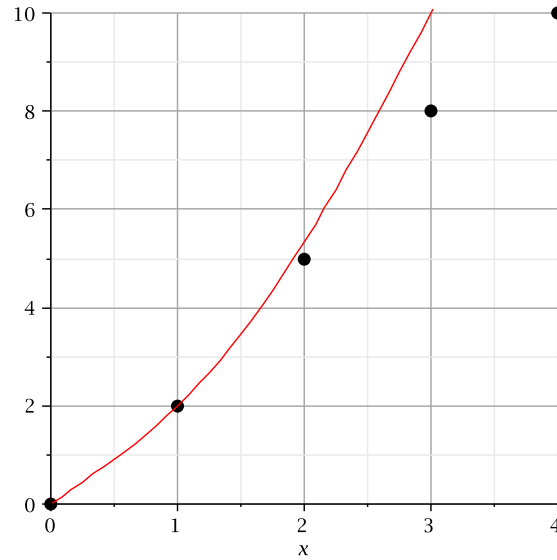&= 0 + 2(x - 0) + 1(x - 0)(x - 2) \\
&= 2x + x(x - 2) = x^2
\end{aligned}
$$

Example

Using a difference table, calculate the fourth order interpolating polynomial using divided differences for the given data.

x_i	0	1	2	3	4
y_i	0	2	5	8	10

Data		Divided Differences			
x_i	$f[x_i]$	$f[x_i, x_{i+1}]$	$f[x_i, x_{i+1}, x_{i+2}]$	$f[x_i, x_{i+1}, x_{i+2}, x_{i+3}]$	$f[x_i, x_{i+1}, x_{i+2}, x_{i+3}, x_{i+4}]$
0	0				
		$\frac{2-0}{1-0} = 2$			
1	2		$\frac{3-2}{2-0} = \frac{1}{2}$		
		$\frac{5-2}{2-1} = 3$		$\frac{0-\frac{1}{2}}{3-0} = -\frac{1}{6}$	
2	5		$\frac{3-3}{3-1} = 0$		$\frac{-\frac{1}{6}-(-\frac{1}{6})}{4-0} = 0$
		$\frac{8-5}{3-2} = 3$		$\frac{-\frac{1}{2}-0}{4-1} = -\frac{1}{6}$	
3	8		$\frac{2-3}{4-2} = -\frac{1}{2}$		
		$\frac{10-8}{4-3} = 2$			
4	10				

$$P_4(x) = 2x + \frac{1}{2}x(x - 1) - \frac{1}{6}x(x - 1)(x - 2)$$

Some notes on Difference Tables

- we can choose to use a lower order polynomial by looking at the divided differences: when the divided differences become small, we can truncate our polynomial there.

- all x_i must be distinct and listed in increasing order

- be careful of x_i that are close together (division by a small number can cause numerical difficulties)

- be careful of errors and irregularities that occur in the data (measurement errors in the data can propagate through the table