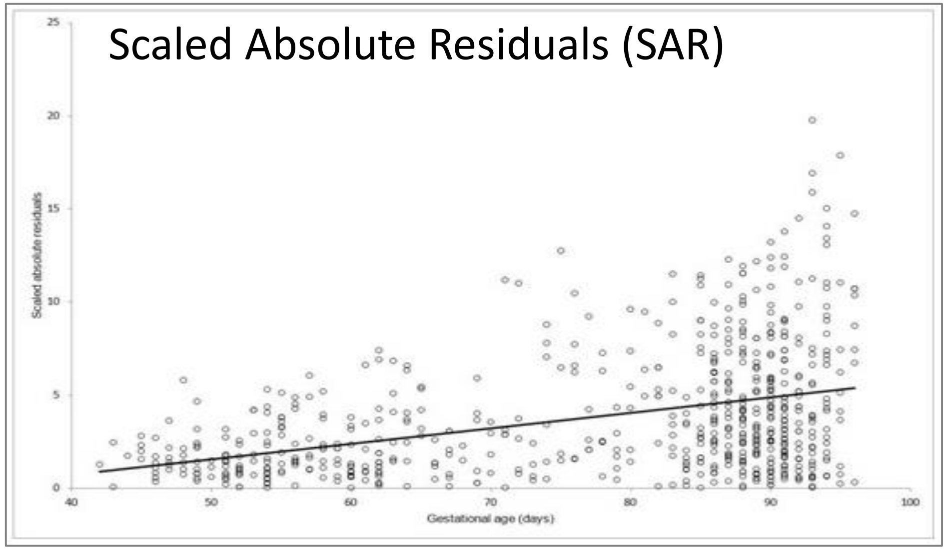
Gradient Descent

A few fine points

Recall this slide



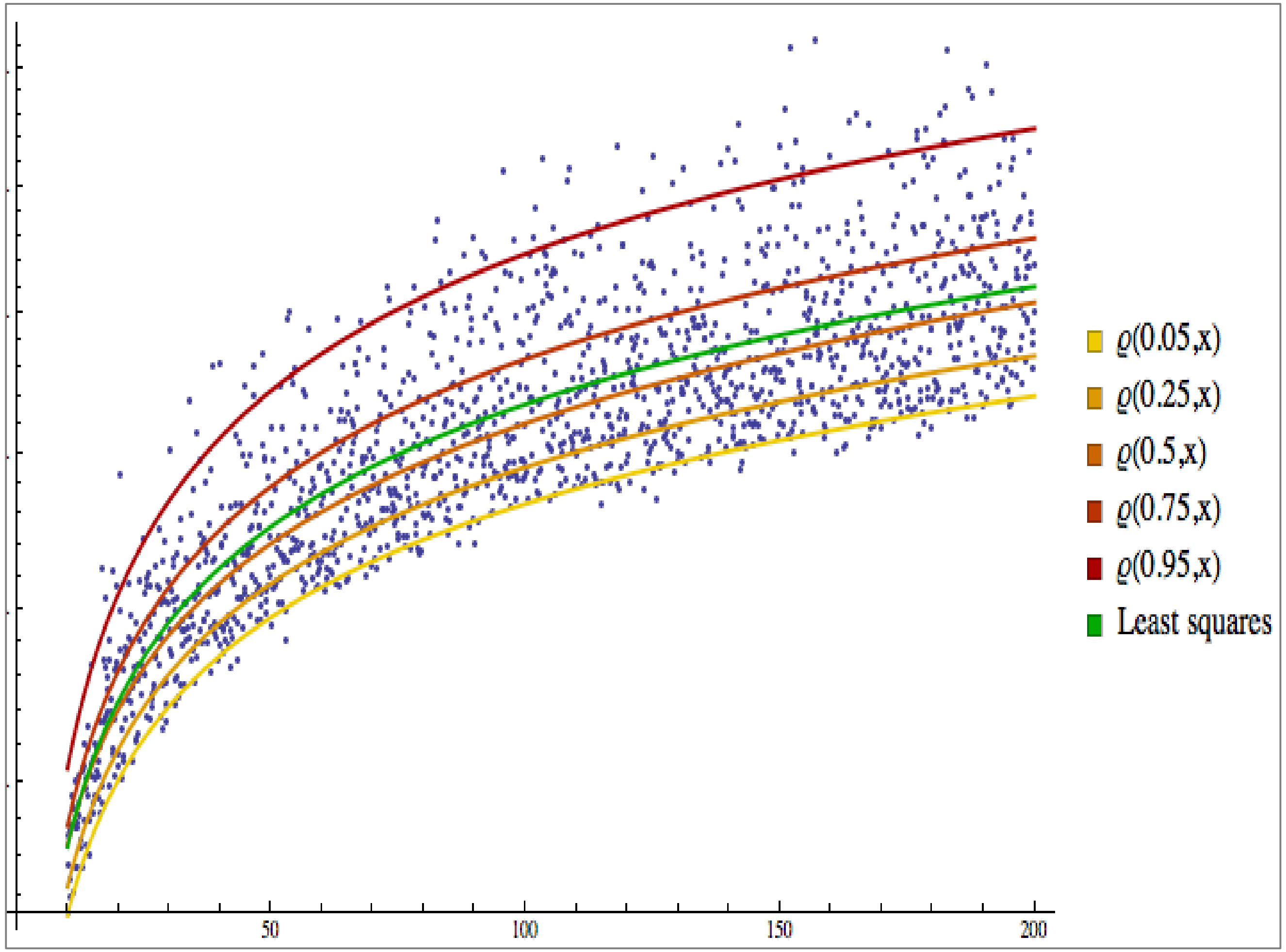
And a little bit more math

$$y(\boldsymbol{a}\,|\boldsymbol{p}, x_{train}) = \boldsymbol{a_0} + \boldsymbol{a_1} * x_1 + \boldsymbol{a_2} * x_2 + \cdots + \boldsymbol{a_n} * x_n$$

$$\varepsilon_p[i] = y_{p,model}[i] - y_{p,observed}[i]$$

$$a = argmin\,\{\boldsymbol{E}[|\varepsilon_p|]\}$$

$$\boldsymbol{E}[|\varepsilon_p|] = \sum_{i=1}^{N} p * \left|\varepsilon_p[i]\,|(\varepsilon_p[i] > 0)]\right| + \sum_{i=1}^{N} (1-p) * \left|[\varepsilon_p[i]|(\varepsilon_p[i] < 0)\right|]$$



Advantages:

Accurate representation of each quantile.

In-depth understanding of underlying behavior.

Disadvantages:

No closed-form solution.

The main idea Behind Gradient Descent

Quantile Regression

$$y(\boldsymbol{a}\,|\boldsymbol{p}, x_{train}) = \boldsymbol{a_0} + \boldsymbol{a_1} * x_1 + \boldsymbol{a_2} * x_2 + \cdots + \boldsymbol{a_n} * x_n$$

$$\varepsilon_p[i] = y_{p,model}[i] \; - y_{p,observed}[i]$$

$$a = argmin\{cost\} = argmin\left\{\boldsymbol{E}\left[|\varepsilon_p|\right]\right\}$$

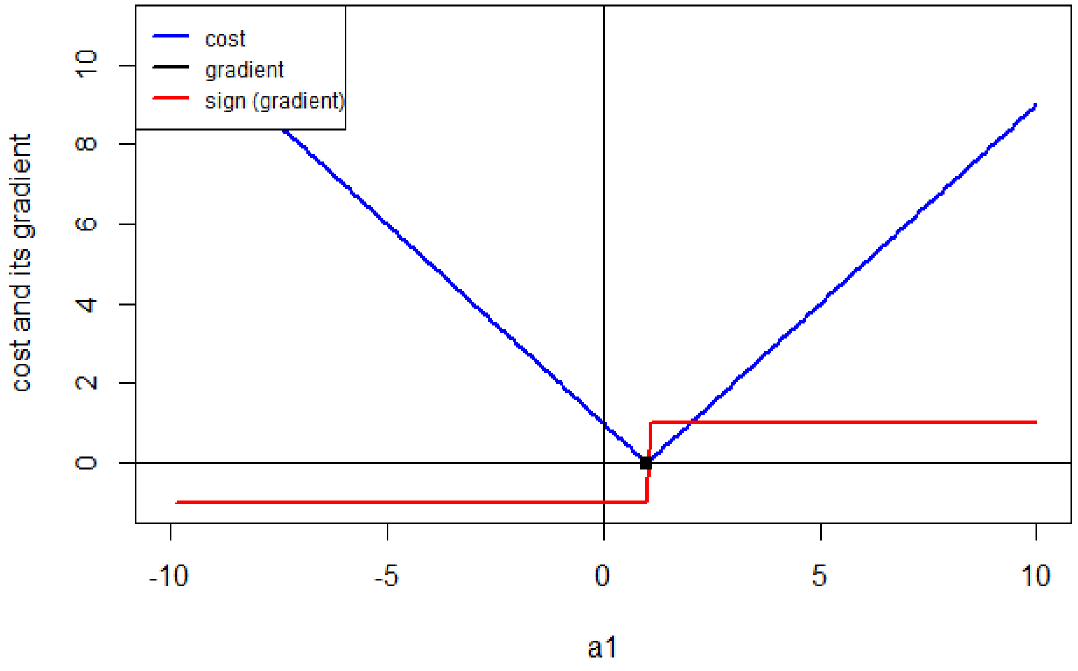
OLS, GLM

$$y(\boldsymbol{a}|x_{train}) = \boldsymbol{a_0} + \boldsymbol{a_1} * x_1 + \boldsymbol{a_2} * x_2 + \cdots + \boldsymbol{a_n} * x_n$$

$$\varepsilon[i] = y_{model}[i] \; - y_{observed}[i]$$

$$a = argmin\{cost\} = argmin\{\boldsymbol{E}[\varepsilon^2]\}$$

$$y(\boldsymbol{a}\,|\boldsymbol{p}, x_{train}) = \boldsymbol{a_1} * x_1$$



$$y(\boldsymbol{a}|x_{train}) = \boldsymbol{a_1} * x_1$$