1) Рассмотрим самый простой случай, когда у нас есть задача бинарной классификации и $\varepsilon_i \in \{0, 1\}$. Алгоритм либо выдает верный ответ с вероятностью p , либо ошибается с вероятностью $1 - p$.

Функция правдоподобия в таком случае:

$$L = \prod_{i=1}^{n} p^{1-\varepsilon_i}(1-p)^{\varepsilon_i}$$

$$lnL = \sum_{i=1}^{n}[(1-\varepsilon_i)\, ln\, p + \varepsilon_i ln\,(1-p)]$$

$$= n\, ln\, p - (ln\, p \sum_{i=1}^{n} \varepsilon_i - ln\,(1-p) \sum_{i=1}^{n} \varepsilon_i) \;\; \rightarrow max$$

$$ln\, \frac{p}{1-p} \sum_{i=1}^{n}(y_i - \; a(x_i)) \;\rightarrow\; min$$

мы получили такую функцию ошибки.

Константный алгоритм при такой функции ошибки будет выдавать все 1 при p > ½ , и 0 при p ≤ ½.

2) Пусть $\varepsilon_i \sim Bi(p, m)$ – биномиальное распределение с параметрами p и m. Для реалистичности ситуации будем подразумевать, что $\varepsilon_i = |y_i - \; a(x_i)|$ так как биномиальное распределение справедливо для положительных величин.
$P(\varepsilon_i) = \binom{m}{\varepsilon_i} p^{\varepsilon_i} q^{m-\varepsilon_i}$

$$L = \prod_{i=1}^{n} \binom{m}{\varepsilon_i} p^{\varepsilon_i}(1-p)^{m-\varepsilon_i}$$

$$lnL = \sum_{i=1}^{n}(ln\frac{m!}{\varepsilon_i!\,(m-\varepsilon_i)!} + \varepsilon_i \ln p + (m - \; \varepsilon_i) ln\,(1-p))$$

$$= n \ln(m!) - \sum_{i=1}^{n} \ln(\varepsilon_i\,!) - \sum_{i=1}^{n} \ln((m-\varepsilon_i)\,!) + \ln p \sum_{i=1}^{n} \varepsilon_i$$

$$+ n\, m \ln(1-p) - ln\,(1-p) \sum_{i=1}^{n} \varepsilon_i) \;\; \rightarrow \begin{matrix} max \\ \varepsilon_i \end{matrix}$$

$$\sum_{i=1}^{n}(ln[\varepsilon_i!\,(m-\varepsilon_i)!] + \varepsilon_i \ln\frac{1-p}{p}) \;\;\rightarrow\;\; min$$

3) Пусть $\varepsilon_i \sim L(\mu, s)$ – логистическое распределение.

$$f(\varepsilon_i, \mu, s) = \frac{e^{-\frac{\varepsilon_i - \mu}{s}}}{s(1 + e^{-\frac{\varepsilon_i - \mu}{s}})^2}$$

Для нашей задачи необходимо, чтобы наиболее вероятное значение было 0, а значит параметр $\mu = 0$.

$$L = \prod_{i=1}^{n} \frac{e^{-\frac{\varepsilon_i}{s}}}{s(1 + e^{-\frac{\varepsilon_i}{s}})^2}$$

$$lnL = \sum_{i=1}^{n} \left[-\frac{\varepsilon_i}{s} - ln\, s - 2\, ln(1 + e^{-\frac{\varepsilon_i}{s}})\right] \to max$$

$$\sum_{i=1}^{n} \left[\ln e^{\frac{\varepsilon_i}{s}} \left(1 + e^{-\frac{\varepsilon_i}{s}}\right)^2\right] \to min$$

Такая функция имеет минимум в точке $\varepsilon_i = 0 \;\; \forall\, i$.

4) $\varepsilon_i \sim$ *Коши(*x_0 *,* γ*)*

$$f(x) = \frac{\gamma}{\pi[\gamma^2 + (x - x_0)^2]}$$

В этом случае аналогично, $x_0 = 0$, чтобы наиболее вероятное значение $\varepsilon_i = 0$

$$L = \frac{\gamma^n}{\pi^n \prod_{i=1}^{n} [\gamma^2 + {\varepsilon_i}^2]}$$

$$ln\, L = n\, ln\, \frac{\gamma}{\pi} - \sum_{i=1}^{n} ln[\gamma^2 + {\varepsilon_i}^2] \to max$$

$$\sum_{i=1}^{n} ln[\gamma^2 + {\varepsilon_i}^2] \to min$$

Получим функцию ошибки, которая по смыслу будет аналогична минимизации сумме квадратов отклонений ответа алгоритма от истинных значений. Добавляется сдвиг и логарифмирование, не влияющее на нашу точку минимума. Также как в случае логистического распределения минимум при $\varepsilon_i = 0 \;\; \forall\, i$.

5) Положим, что $\varepsilon_i = |y_i - a(x_i)|$ для $\varepsilon_i \sim exp(\lambda)$

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 \text{ иначе} & \end{cases}$$

$$L = \begin{cases} \lambda^n \, e^{-\lambda \sum_{i=1}^{n} \varepsilon_i}, & \sum_{i=1}^{n} \varepsilon_i \geq 0 \\ 0, & \text{иначе} \end{cases}$$

Вторую часть можем опустить по предположению выше. И получаем:

$$lnL = n \ln \lambda - \lambda \sum_{i=1}^{n} \varepsilon_i \ \rightarrow max$$

$$\sum_{i=1}^{n} |y_i - a(x_i)| \ \rightarrow \ min$$

В отсутствии предположения алгоритм минимизировал бы просто сумму отклонений и при настройке алгоритма с такой функцией ошибки «оптимальным» $a(x_i) = +\infty$.