Funkcja kosztu. Chcemy wybrać taki punkt $v \in \mathbb{R}^D$, żeby błąd zastąpienia wszystkich N elementów zbioru X przez v był najmniejszy. Pokażemy, że wynik zależy od tego jak zdefiniujemy błąd. Rozpatrzymy trzy możliwości:

$$\begin{aligned}
\mathrm{ME}(X; v) &= \tfrac{1}{N} \textstyle\sum_i \|x_i - v\|_2 \\
&\text{mean error - sumaryczny błąd,} \\
\mathrm{MSE}(X; v) &= \tfrac{1}{N} \textstyle\sum_i \|x_i - v\|_2^2 \\
&\text{mean squared-error - błąd kwadratowy,} \\
\mathrm{MaxE}(X; v) &= \max_i \|x_i - v\|_2 \\
&\text{maximal error - największy błąd,}
\end{aligned}$$

gdzie $\|w\|_2 = \sqrt{\sum_{i=1}^D w_i^2}$ oznacza normę euklidesową. Oczywiście zmiana sumy błędów po punktach na średnią nie powoduje zmiany rozwiązania, często obie wielkości są stosowane wymiennie. Punkt, który minimalizuje zadane kryterium jest rozwiązaniem problemu optymalizacyjnego. *Funkcja kosztu* jest podstawą definicji modelu uczenia maszynowego, i oznacza funkcję której minimalizacja prowadzi do rozwiązania interesującego nas zadania.

Minimalizacja gradientowa w minimalizacji błędów. Pokażemy teraz w jaki sposób minimalizować przytoczone na początku rozdziału funkcje błędów za pomocą metod gradientowych. Przez h oznaczamy wielkość kroku.

Rozpatrzmy najpierw przypadek minimalizacji sumarycznego błędu ME. Gradient funkcji ME to

$$\nabla \mathrm{ME}(X; v) = \frac{1}{N} \sum_i \frac{v - x_i}{\|v - x_i\|}.$$

Oznacza to, że jeżeli w danej iteracji znajdujemy się w punkcie v_k, to następny krok jest dany przez (minus ze składnika $v - x_i$ został wyciągnięty przed ułamek):

$$v_{k+1} = v_k + h \frac{1}{N} \sum_i \frac{x_i - v_k}{\|x_i - v_k\|}.$$

Geometrycznie oznacza to, że jeżeli środek układu współrzędnych umieścimy w punkcie v_k, to patrzymy z tego punktu w kierunku każdego punktu ze zbioru danych, i poruszamy się w stronę uśrednionego kierunku. Ponadto, minimum zostanie osiągnięte, gdy gradient się wyzeruje, czyli gdy średni kierunek jest zerowy.

Teraz przejdziemy do przypadku funkcji MSE. Gradient funkcji błędu kwadratowego:

$$\nabla \mathrm{MSE}(X; v) = 2(v - \mathrm{mean}(X)).$$

Warto zauważyć, że on zależy tylko od średniej. Przeprowadzając analogiczną analizę jako poprzednio, jeżeli znajdujemy się w punkcie v_k, to w następnym kroku znajdziemy się w

$$v_{k+1} = v_k + 2h(\mathrm{mean}(X) - v_k).$$

Czyli jeżeli znajdujemy się w punkcie v_k, to poruszmy się w kierunku $\mathrm{mean}(X)$.

Na koniec zajmijmy się funkcją MaxE. Połóżmy $f(v) = \mathrm{MaxE}(X, v)$. Oczywiście mamy

$$f(v) = \mathrm{MaxE}(X, v) = \max_i \|v - x_i\|.$$

W k-tym kroku znajdujemy się w punkcie v_k. Możemy teraz założyć (przeciwny przypadek jest w praktyce pomijalny), że istnieje dokładnie jeden punkt $x_j \in X$, który jest najdalej od v_k. Wtedy

$$f(v) = \|v - x_j\|, \text{ dla } v \text{ odpowiednio bliskich } v_k.$$

W konsekwencji otrzymujemy, że

$$\nabla f(v) = \frac{v - x_j}{\|v - x_j\|}, \text{ gdzie } j = \operatorname*{argmax}_i\{\|v - x_i\|\}$$