Modele statystyczne, nadzorowane nauczanie i przybliżenie funkcji

Model statystyczny dla wspólnego rozkładu Pr(X,Y)

$$Y = f(X) + \varepsilon$$

$$\mathrm{E}(\varepsilon) = 0$$

$$f(x) = \mathrm{E}(Y|X = x)$$

Uczenie nadzorowane

$$Y = f(X) + \varepsilon$$

$$\mathcal{T} = (x_i, y_i),\ i = 1, \ldots, N$$

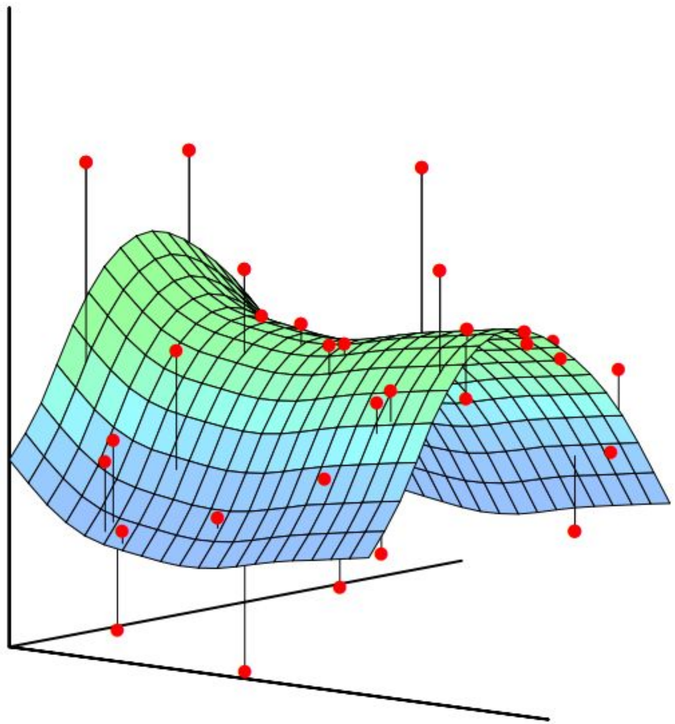
Przybliżanie funkcji

- Pary danych $\{x_i, y_i\}$ utożsamiamy jako punkty w (p + 1)-wymiarowej przestrzeni euklidesowej.
- Przyjmujemy dziedzinę $\mathbb{R}^p$.
- Liniowe rozwinięcia bazowe $f_\theta(x) = \sum_{k=1}^{K} h_k(x)\theta_k,$
- Transformacja sigmoidalna wspólna dla modeli sieci neuronowych

$$h_k(x) = \frac{1}{1 + \exp(-x^T\beta_k)}$$

- Inne $x_1^2,\ x_1x_2^2,\ \cos(x_1)$

- Metoda najmniejszych kwadratów



$$\mathrm{RSS}(\theta) = \sum_{i=1}^{N} (y_i - f_\theta(x_i))^2$$

- Estymacja największego prawdopodobieństwa

Próba $y_i, \; i = 1, \ldots, N$ o rozkładzie $\Pr_\theta(y)$

Prawdopodobieństwo logarytmiczne $L(\theta) = \sum_{i=1}^{N} \log \Pr_\theta(y_i)$

$$\Pr(Y|X, \theta) = N(f_\theta(X), \sigma^2)$$

$$L(\theta) = -\frac{N}{2} \log(2\pi) - N \log \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^{N} (y_i - f_\theta(x_i))^2$$

- Wielomianowe prawdopodobieństwo funkcji regresji Pr(G|X) dla wyjścia jakościowego G

$$\Pr(G = \mathcal{G}_k | X = x) = p_{k,\theta}(x),\ k = 1, \ldots, K$$

$$L(\theta) = \sum_{i=1}^{N} \log p_{g_i,\theta}(x_i)$$

Modele regresji strukturalnej

Problematyka

$$\mathrm{RSS}(f) = \sum_{i=1}^{N} (y_i - f(x_i))^2$$

$$x_i, y_{i\ell}, \; \ell = 1, \ldots, N_i$$