GT Deep Learning: 02

Vincent Perrollaz

Institut Denis Poisson

04 Février 2026

Linear Regression?

Probabilistic Relaxation in 1d

Probabilistic Relaxation in 2d

Linear Regression

Data

$$(x_k, y_k)_{1 \leq k \leq N} \in (\mathbb{R}^2)^N \tag{1}$$

Goal

$$\arg\min_{a,b \in \mathbb{R}^2} \sum_{k=1}^{N} (ax_k + b - y_k)^2 \tag{2}$$

Questions

1. Why linear functions?
2. Why sum of squares?

Error : $e \mapsto e^2$ vs $e \mapsto |e|$

Data $(x_k)_{1 \leq k \leq \mathbb{R}} \in \mathbb{R}^N$

Mean $\arg\min_{x \in \mathbb{R}} \sum_{k=1}^{N} (x - x_k)^2$

Median $\arg\min_{x \in \mathbb{R}} \sum_{k=1}^{N} |x - x_k|$

Warning outliers $\Longrightarrow$ median $>$ mean

Historically computation easier with squares

To be Linear or not to be linear?

Data $(x_k, y_k)_{1 \leq k \leq N} \in (\mathbb{R}^2)^N$

Generalized Linear

$$\arg\min_{a \in \mathbb{R}^d} \sum_{k=1}^{N} \left(\sum_{i=1}^{d} a_i g_i(x_k) - y_k \right)^2$$

Demo in companion notebook

Limits Inverse Problem in companion notebook

Linear Regression?

Probabilistic Relaxation in 1d

Probabilistic Relaxation in 2d

Data $(x_k)_{1\leq k\leq N} \in \mathbb{R}^N$

Modelling $X_1, \ldots, X_N, M, S$ r.v.

$$\begin{aligned}
\mathbb{P}(X_1 = x_1, \ldots, X_N = x_n \mid S = \sigma, M = m) &= \prod_{k=1}^{N} \mathbb{P}(X_k = x_k \mid S = \sigma, M = m) \\
&= \prod_{k=1}^{N} \frac{\exp(-(x_k - m)^2/2\sigma)}{\sqrt{2\pi\sigma}} \\
&= \frac{\exp(-\sum_{k=1}^{N}(x_k - m)^2/2\sigma)}{(2\pi\sigma)^{N/2}}
\end{aligned}$$

Bayes Theorem

Since $\mathbb{P}(A \mid B) := \mathbb{P}(A \cap B)/\mathbb{P}(B)$

Then $\boxed{\mathbb{P}(A \mid B) = \mathbb{P}(B \mid A)\mathbb{P}(A)/\mathbb{P}(B)}$

Taking the logarithm

$$\ln(\mathbb{P}(A \mid B)) = \ln(\mathbb{P}(B \mid A)) + \ln(\mathbb{P}(A)) - \ln(\mathbb{P}(B))$$

Maximum likelihood indifferent $(A_i)_{1 \leq i \leq d} + B \implies$

$$\arg\max_{1 \leq i \leq d} \ln(\mathbb{P}(B \mid A_i)).$$

Previous situation

$$\arg\max_{m,\sigma} \mathbb{P}(S = \sigma, M = m \mid X_1 = x_1, \ldots, X_N = x_n)$$

Becomes

$$\arg\max_{m,\sigma} \; -\frac{\sum_{k=1}^{N}(x_k - m)^2}{2\sigma} - \frac{\ln(2\pi\sigma)}{2N}$$

So

$$m = \bar{x} := \frac{1}{N}\sum_{k=1}^{N} x_k \quad \text{and} \quad \sigma = \frac{1}{N}\sum_{k=1}^{N}(x_k - \bar{x})^2.$$

Why the Normal Distribution?

- Fixed mean and variance on $\mathbb{R}$: normal distribution maximizes entropy
- Fixed mean on $\mathbb{R}^+$: exponential distribution maximizes entropy
- On $[a, b]$: uniform distribution maximizes entropy
- More assumptions $\Longrightarrow$ other distributions

Inference/Decision split

Generative AI distribution $\Longrightarrow$ sampling!

Bayesian Way

$$\begin{aligned}
&\mathbb{P}(X = x \mid X_1 = x_1, \ldots, X_N = x_n) \\
&= \int \mathbb{P}(X = x, S = \sigma, M = m \mid X_1 = x_1, \ldots) d\sigma dm \\
&= \int \mathbb{P}(X = x \mid S = \sigma, M = m) \\
&\qquad \mathbb{P}(S = \sigma, M = m \mid X_1 = x_1, \ldots) d\sigma dm.
\end{aligned}$$

Linear Regression?

Probabilistic Relaxation in 1d

Probabilistic Relaxation in 2d

Linear Regression Bis

Data $(x_k, y_k)_{1 \leq k \leq N} \in (\mathbb{R}^2)^N$

Modelling $\mathbb{P}(y \mid x, a, b) := \frac{e^{-\frac{(y-ax-b)^2}{2\sigma}}}{\sqrt{2\pi\sigma}}$

Meaning $X_1, \ldots, X_N, Y_1, \ldots, Y_N, A, B$ r.v.

$$\mathbb{P}(Y = y \mid X = x, A = a, B = b)$$

$+$ Disintegration Theorem

Same Computation

- Conditional Independance
- Maximum Likelihood
- We end up with

$$\arg\max_{a,b} \; -\frac{\sum_{k=1}^{N}(y_k - ax_k - b)^2}{2\sigma} + \frac{N}{2}\ln(2\pi\sigma)$$

Inverse Problem

Gaussian Mixture Distribution see notebook

Maximum Likelihood numerical optimization!

General

1. Neural Network Output: distribution parameters
2. Maximum Likelihood: Loss function to optimize

Exercise

1. Compute the loss functions for the inverse problem and a neural network.
2. Apply the same ideas to a classification problem.