Deepest Regression Estimator

Zayaan and Ananya

University of Waterloo

2024

Some background on linear regression

Simple problem statement

Given some set of points $S = \{(x_1, y_1), ..., (x_n, y_n)\}$, where each $x_i \in \mathbb{R}^k$, $y_i \in \mathbb{R}$, we aim to fit a hyperplane of the form $Y = X^T B + \epsilon$, where $X = (x_1, ..., x_n)^t$, $B = (b_1, ..., b_n)^t \in \mathbb{R}^n$, and $\epsilon \in \mathbb{R}$.

Least squares estimator

Given a loss function $l(B, \epsilon) = \frac{1}{2}(Y_i - y_i)^2$, we want to find a B such that we minimize the average loss on the sample set of points.
i.e, we solve for $argmin_{B,\epsilon} \frac{1}{n} \sum l_i(B, \epsilon) = argmin_{B,\epsilon} \frac{1}{n} \sum \frac{1}{2}(B^T X_i + \epsilon - y_i)^2$.
Encoding ϵ into B, this problem is the equivalent of finding $argmin_B ||y - X^T B||^2$.

Notions of Robustness

- Outliers, Non-normal error distribution, Heteroscedasticity

Regression Depth

Definition

First consider the residuals of S relative to a fit B, $r_i(B) = Y_i - (B^T X_i + \epsilon)$. Then, $rdepth(B, S) = min\{\#(r_i(B) \geq 0 \text{ and } X_i^T u < v) + \#(r_i(B) \leq 0 \text{ and } X_i^T u > v)\}$, where $u \in \mathbb{R}^k$ is over all unit vectors, and $v \in \mathbb{R}$, such that $X_i^T u \neq v, X_i \in S$.

Interpretation

- Smallest number of observations needed to pass when tilting B till it becomes vertical
- Smallest number of observations needed to be removed to make B a nonfit

Remarks

- Measures balance of dataset about the linear fit
- $0 \leq rdepth(B, S) \leq n$

Implementing Regression Depth

Naive algorithm

This has time complexity $o(n^{k-1}log(n))$, as we have to do more or less a brute force calculation as to how many positive and negative residuals lie on each side of every possible hyperplane, through every possible point in the data set. We can do better.

Fast Approximation algorithm

1. Set all $(x_i, y_i) \to (x_i, r_i)$ $o(n)$
2. repeat m times: $o(m)$
   1. generate random $k - subset$ of $n, \{i_1, ..., i_k\}$ $o(k)$
   2. project all datapoints $\{x_{i_1}, ..., x_{i_k}\}$ on to a vertical plane P through some line L with direction perpecidular to all datapoints $o(k^3 + nk)$
   3. compute bivariate regression depth of L in P and keep track of this minimum $o(nlogn)$

This gives total runtime of $o(mnk + mk^3 + mnlogn)$

Deepest Regression Estimator

Definition

$DR(S) = argmax_B rdepth(B, S)$
i.e the fit with maximal depth

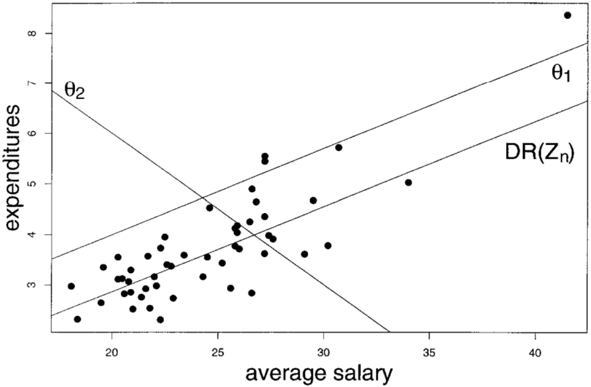
Breakdown Value

1. Defines the smallest percentage of 'contamination' data that can be added to S such that B becomes useless.
2. In any dimension, this evaluates to at most $\frac{1}{3}$ for the deepest regression estimator.
3. This value is $\frac{1}{n}$ for ordinary least squares.

Algorithm

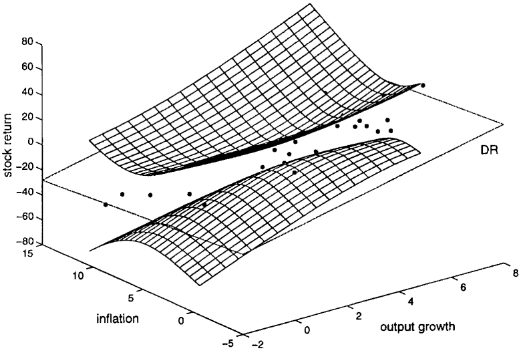
MEDSWEEP

Sample Experiment



Here the data, with n=51 has observations in 2 dimensions. The lines θ_1, θ_2 are generated both with depth 2. The he line generated by the deepest regression estimator, with depth 23, shows a significantly better fit.

Experiment



Here we see that with 1000 samples of the stock return dataset, we form an accurate deepest regression plane, with the upper and lower surfaces representing the 95% confidence regions.

VAE - Metric Learning Model

Goal

To develop architecture that learns transformation invariant metric on any set of data. i.e it should learn a continuous representation of the input.

Architecture

Encoder

Our modified VAE takes into the encoder:

1. Vector
2. Transformed (not id) vector
3. Different pair vector
4. Specified Metric

This creates latent representations of each, namely z, z_r, z_b, m

Architecture

Loss functions

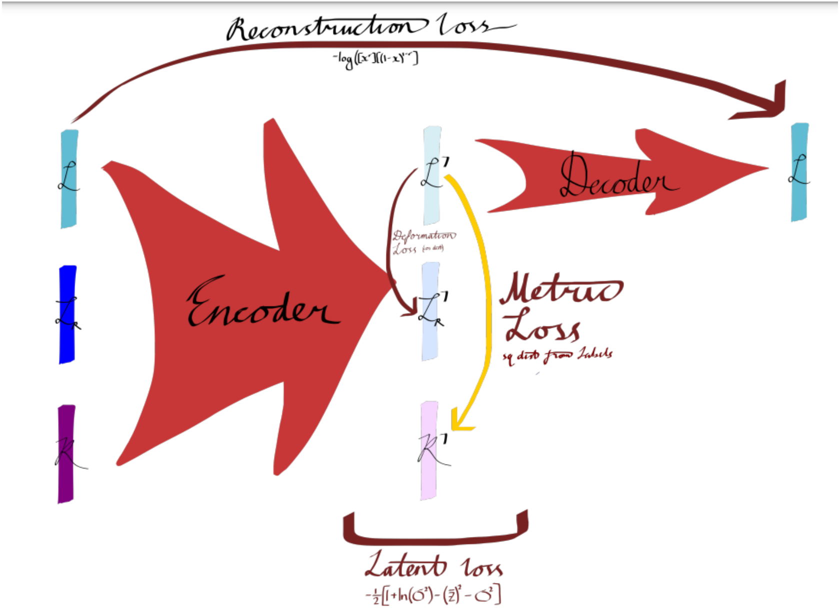
We modify the loss function of the standard VAE to add in 2 more losses to learn:

1. transformation invariant representation of data
2. morphs latent space into one with defined metric

This gives us the following loss function:

1. $D_{KL}(q_\psi(z|x)||p_\psi(z)) - \mathbb{E}_{z\ q_\psi}[log(p_\psi(x|z))] + \frac{10}{n}\sum_{i=1}^{n}(z - z_r)^2$
2. $\frac{||z||\cdot||z_b||}{z\cdot z_b} - m$

Diagram


Reconstruction Loss
-log([x^y][(1-x)^{1-y}])
Encoder
Decoder
Deformation Loss
Metric Loss
sq dist from labels
Latent loss
-½[1+ln(σ²)-(z̄)²-σ²]

Experiments-Max Overlap Problem

Based on the shape overlap dataset, our VAE is able to predict the correct max overlap of 1783 pixels my creating a continuous representation of the shapes.

Learning rates with two optimizers were 1e-8 and 1e-10 with fully connected encoder network containing 4 layers of 4000, 500x3 neurons and the decoder network containing 2x500 neuron layers.

References

1. Ishfaq, Haque, and Ruishan Liu. TVAE: Deep Metric Learning Approach for Variational Autoencoder. 2017.
2. Rousseeuw, Peter J., and Mia Hubert. "Regression Depth." Journal of the American Statistical Association, vol. 94, no. 446, June 1999, pp. 388–402, https://doi.org/10.1080/01621459.1999.10474129. Accessed 9 Oct. 2020.
3. Stromberg, Arnold J., and David Ruppert. "Breakdown in Nonlinear Regression." Journal of the American Statistical Association, vol. 87, no. 420, Dec. 1992, p. 991, conservancy.umn.edu/server/api/core/bitstreams/40f3aced-0109-450d-9fe2-d6da7c918070/content, https://doi.org/10.2307/2290636. Accessed 31 July 2024.
4. Van Aelst, Stefan, et al. "The Deepest Regression Method." Journal of Multivariate Analysis, vol. 81, no. 1, Apr. 2002, pp. 138–166, https://doi.org/10.1006/jmva.2001.1997. Accessed 9 Oct. 2020.
5. Van Aelst, Stefan, and Peter J Rousseeuw. "Robustness of Deepest Regression." Journal of Multivariate Analysis, vol. 73, no. 1, Apr.