1 Regression

1.1 Terminology

- Data consists of **pairs** $(\mathbf{x}_n, y_n)$, where y_n is the n'th output and x_n is a vector of D inputs. The number of pairs N is the data-size and D is the dimensionality.

- Two goals of regression: **prediction** and **interpretation**

- The regression function: $y_n \approx f_w(\mathbf{x}_n) \; \forall n$

- Regression finds correlation not a causal relationship.

- **Input variables** a.k.a. covariates, independent variables, explanatory variables, exogenous variables, predictors, regressors.

- **Output variables** a.k.a. target, label, response, outcome, dependent variable, endogenous variables, measured variable, regressands.

1.2 Linear Regression

- Assumes linear relationship between inputs and output.

- $y_n \approx f(\mathbf{x}_n) := w_0 + w_1 x_{n1} + ... + w_D x_{nD}$
$:= \tilde{\mathbf{x}}_n^T \tilde{\mathbf{w}}$ contain the additional offset term (a.k.a. bias).

- Given data we learn the weights $\mathbf{w}$ (a.k.a. estimate or fit the model)

- Overparameterisation $D > N$ eg. univariate linear regression with a single data point $y_1 \approx w_0 + w_1 x_{11}$. This makes the task under-determined (no unique solution).

1.3 Loss Functions $\mathcal{L}$

- A loss function (a.k.a. energy, cost, training objective) quantifies how well the model does (how costly its mistakes are).

- $y \in \mathbb{R} \Rightarrow$ desirable for cost to be symmetric around 0 since $\pm$ errors should be penalized equally.

- Cost function should penalize "large" mistakes and "very large" mistakes similarly to be robust to outliers.

- Mean Squared Error:
$\mathsf{MSE}(\mathbf{w}) := \frac{1}{N} \sum_{n=1}^{N} [y_n - f_\mathbf{w}(\mathbf{x}_n)]^2$
not robust to outliars.

- Mean Absolute Error:
$\mathsf{MAE}(\mathbf{w}) := \frac{1}{N} \sum_{n=1}^{N} |y_n - f_\mathbf{w}(\mathbf{x}_n)|$

- Convexity: a function is convex iff a line segment between two points on the function's graph always lies above the function.

- Convexity: a function $h(\mathbf{u}), \mathbf{u} \in \mathbb{R}^D$ is convex if $\forall\, \mathbf{u}, \mathbf{v} \in \mathbb{R}^D, 0 \leq \lambda \leq 1$:
$h(\lambda\mathbf{u} + (1-\lambda)\mathbf{v}) \leq \lambda h(\mathbf{u}) + (1-\lambda)h(\mathbf{v})$
Stirctly convex if $\leq \Rightarrow <$

- Convexity, a desired computational property: A strictly convex function has a unique global minimum $\mathbf{w}^*$. For convex functions, every local minimum is a global minimum.

- Sums of convex functions are also convex $\Rightarrow$ MSE combined with a linear model is convex in $\mathbf{w}$.

- Proof of convexity for MAE:
$\mathsf{MAE}(\mathbf{w}) := \frac{1}{N} \sum_{n=1}^{N} \mathcal{L}_n(\mathbf{w}), \mathcal{L}_n(\mathbf{w}) = |y_n - f_\mathbf{w}(\mathbf{x}_n)|$
$\mathcal{L}_n(\lambda w_1 + (1-\lambda)w_2) \leq \lambda\mathcal{L}_n(w_1) + (1-\lambda)\mathcal{L}_n(w_2)$
$|y_n - x_n^T(\lambda w_1 + (1-\lambda)w_2)| \leq \lambda|y_n - x_n^T w_1| + (1-\lambda)|y_n - x_n^T w_2|$
$(1-\lambda) \geq 0 \Rightarrow (1-\lambda)|y_n - x_n^T w_2| = |(1-\lambda)y_n - (1-\lambda)x_n^T w_2|$
$a = \lambda y_n - \lambda x_n^T w_1, b = (1-\lambda)y_n - (1-\lambda)x_n^T w_2$
$a + b = y_n - x_n^T(\lambda w_1 + (1-\lambda)w_2)$
$|a+b| \leq |a| + |b| \Rightarrow \mathcal{L}_n(\mathbf{w})$ convex $\Rightarrow \mathsf{MAE}(\mathbf{w})$ convex

- Huber loss:
$$\mathcal{H}uber(e) := \begin{cases} \frac{1}{2}e^2 & \text{, if } |e| \leq \delta \\ \delta|e| - \frac{1}{2}\delta^2 & \text{, if } |e| > \delta \end{cases} \text{ convex,}$$
differentiable, and robust to outliers but setting δ is not easy.

- Tukey's bisquare loss:
$$\frac{\partial\mathcal{L}}{\partial e} := \begin{cases} e\{1 - e^2/\delta^2\}^2 & \text{, if } |e| \leq \delta \\ 0 & \text{, if } |e| > \delta \end{cases} \text{ non-convex,}$$
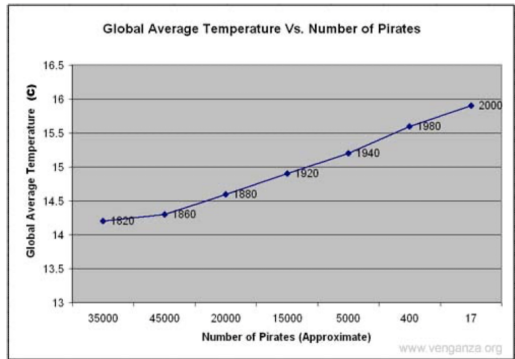but robust to outliers.
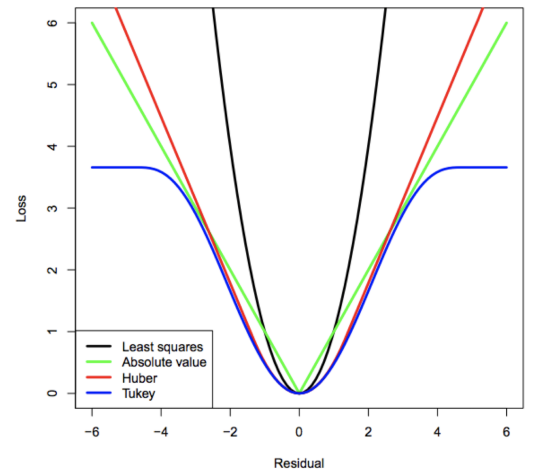
2 Optimisation

- Given $\mathcal{L}(\mathbf{w})$ we want $\mathbf{w}^* \in \mathbb{R}^D$ which minimises the cost: $\min_\mathbf{w} \mathcal{L}(\mathbf{w}) \rightarrow$ formulated as an optimisation problem

- Local minimum $\mathbf{w}^* \Rightarrow \exists \epsilon > 0$ s.t.
$\mathcal{L}(\mathbf{w}^*) \leq \mathcal{L}(\mathbf{w}) \; \forall \mathbf{w}$ with $\|\mathbf{w} - \mathbf{w}^*\| < \epsilon$

- Global minimum $\mathbf{w}^*$, $\mathcal{L}(\mathbf{w}^*) \leq \mathcal{L}(\mathbf{w}) \; \forall \mathbf{w} \in \mathbb{R}^D$

2.1 Smooth Optimisation

3 This is RGB red text.

For $x \in [r_{i-1}, r_i]$
$r(x) = \tilde{a}_1 x + \tilde{b}_1 + \sum_{j=2}^{m} \tilde{a}_j (x - \tilde{b}_j)_+$

For $x \in [r_{i-1}, r_i]$
$r(x) = \tilde{a}_1 x + \tilde{b}_1 + \sum_{j=2}^{m} \tilde{a}_j (x - \tilde{b}_j)_+$

Global Average Temperature Vs. Number of Pirates

A line chart with y-axis "Global Average Temperature (C)" ranging from 13 to 16.5, and x-axis "Number of Pirates (Approximate)" with values 35000, 45000, 20000, 15000, 5000, 400, 17. Data points labeled 1820, 1860, 1880, 1920, 1940, 1980, 2000 show an increasing trend. www.venganza.org

Once upon a time in the whimsical town of Serendipity Springs, there lived a peculiar inventor named Professor Quixotic. Known for his wild ideas and even wilder inventions, the professor spent his days tinkering away in his eccentric laboratory.

. . . .

One day, while experimenting with a contraption designed to turn laughter into energy, Professor Quixotic accidentally created a portal to a parallel universe. As the portal crackled with energy, it sucked the professor and his loyal robotic sidekick, Gizmo, into a world where everything was upside down.

In this topsy-turvy land, gravity worked in reverse, causing the duo to float above the ground. Trees grew with their roots in the air, and people walked on the ceilings of their homes. It was a place of perpetual confusion and constant laughter.

As Professor Quixotic and Gizmo explored this strange realm, they encountered a group of upside-down creatures called the Giggleglides. These whimsical beings thrived on laughter and had the power to make anything levitate with a simple joke.

Realizing the potential of this upside-down world, Professor Quixotic teamed up with the Giggleglides to create a laughter-powered flying machine. With the combined ingenuity of the professor and the Giggleglides' comedic talents, they built the most extraordinary aircraft the upside-down world had ever seen.

The laughter-powered flying machine soared through the topsy-turvy skies, spreading joy and amusement wherever it went. The upside-down inhabitants were thrilled to have a new form of transportation that brought them even closer to the upside-down stars.

Eventually, Professor Quixotic and Gizmo found another portal, one that led them back to Serendipity Springs. The laughter-powered flying machine became a symbol of the magical adventure, and the town embraced the upside-down spirit with open arms.

And so, the town of Serendipity Springs became a place where laughter powered not only inventions but also the hearts of its inhabitants. Professor Quixotic continued his quirky experiments, inspired by the upside-down world, and the memory of the laughter-filled adventure remained etched in the hearts of all who experienced it.