

04 - Linear Regression

Ethan Graham

20th June 2023

Reminder: Supervised Classification

We want to minimize: $E(\mathbf{w}) = \sum_{n=1}^N L(y(\mathbf{x}_n; \mathbf{w}), t_n)$ Where we have: - x : feature vector - w : model parameters - t : label - y : predictor - L loss function - E error function

ML is an optimization problem!

Line parametrisation

Mathematically we express a ($2D$) line as $y(x; \mathbf{w}) = w_0 + w_1 \cdot x$ Given N pairs $\{x_i, t_i\}$, we want to find the line that most closely fits the observations. Essentially, assuming D dimensional space, we are looking for optimal line parameters w_0, \dots, w_D Once again, the natural measure of distance is Euclidean distance. In practice we mostly use the **squared euclidean distance** however. This penalizes greater distances more harshly. This also allows us to express the problem as a **least-squares** problem *recall Nummet*

Once we have found optimal parameters $\mathbf{w} = w_{\star_0}, \dots, w_{\star_D}$, we predict with the

formula: $y_t = w_0 + \mathbf{w}_{1 \rightarrow D} \cdot \mathbf{x}_t$ Or $y_t = \mathbf{w}^T \begin{bmatrix} 1 \\ x_0 \\ \dots \\ x_D \end{bmatrix}$

In D -dimensional space, we no longer try to fit a line but a **hyperplane** to our dataset.

Because the output is still one dimensional, we can use the least-squares formulation from previously $\min_w \frac{1}{N} \sum_{i=1}^N (\mathbf{w}^T \cdot \mathbf{x}_i - t_i)^2$

A few short notes on optimization

There are a few methods that we commonly use for optimizing a minimization problem such as this - Gradient Descent (*applicable to many problems*) - Closed

Form Solution (*only really applicable to linear regression*)

Linear Regression Closed-Form Solution

We want to find a solution $\nabla_w E(w) = 0$ implying that we have found a minimum. Expansion yields the following formula for finding the optimal line parameters w^\star $w^\star = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{t}$. This is just the least-squares solution $\mathbf{X}^T \mathbf{t} = \mathbf{X}^T \mathbf{X} w^\star$

Evaluation Metrics for Regression

One of the common evaluation metrics is the **mean-squared error**

$$MSE = \frac{1}{N_t} \sum_{i=1}^{N_t} (y_i - t_i)^2$$

Another common metric is the **root mean-squared error** where we simply take the square root of MSE.

We also use **mean absolute error**

$$MAE = \frac{1}{N_t} \sum_{i=1}^{N_t} |y_i - t_i|$$

or **mean absolute percentage error**

$$MAPE = \frac{1}{N_t} \sum_{i=1}^{N_t} \left| \frac{y_i - t_i}{t_i} \right|$$

Taking percentage with respect to the true value may be easier to interpret.

Final notes on interpreting a linear model

It is important to be careful of different magnitudes in the features since it may lead to incorrect predictions (*some features overpower others*). We can address this by normalizing the dataset.