



Social Statistics - Away Day

Dr Diego Perez Ruiz

May / 2024

Further Work

Read the book "The Elements of Statistical Learning" by Tibshirani, Hastad, and Tibshirani. It is a great resource for understanding the theoretical foundations of machine learning.

(Short) Discussion

Mutual information is an important criterion for feature selection, equivalent to minimising the conditional target entropy $H(Y|X)$, which corresponds to minimising the MSE.

Q&A

Thanks! Any questions?



Decoding Data Quality: An Introduction to Entropy and Mean Square Error (MSE).

Entropy and the Mean Square Error (MSE) are two fundamental concepts in the field of data science. Entropy is a measure of the uncertainty or randomness in a system, while MSE is a measure of the average squared difference between the estimated values and the true values.

A Mathematical Theory of Communication By C. E. SHANNON

What is Mutual Information?

Mutual information is a measure of the dependence between two variables. It is defined as the reduction in uncertainty about one variable given knowledge of the other variable.

What is Entropy?

In simple terms, entropy measures the uncertainty or randomness in a set of possible outcomes.



Entropy is a measure of the uncertainty or randomness in a system. It is defined as the negative logarithm of the probability of an event occurring.

Regression

Regression is a statistical method used to model the relationship between a dependent variable and one or more independent variables.

The Normal Case

The normal case is a common scenario in regression analysis where the data points are distributed around a central value, forming a bell-shaped curve.

Adaptive Regression





Social Statistics - Away Day

Dr Diego Perez Ruiz

May / 2024



Decoding Data Quality:
An introduction to Entropy and Mean Square Error (MSE).

A Mathematical Theory of Communication By C. E. SHANNON

- Define what the Mutual Information is.
- Interpretation of the MSE can be interpreted as the expected variance of the estimation error.
- Mutual information can be thought of as the feature selection using when the conditional distribution of the estimation error is Gaussian.

What is Mutual Information?

$$I(X; Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$$
$$I(X; Y) = \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}$$

The Normal Case

$$W(x) = \frac{1}{\sigma \sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2}\right)$$
$$f(x) = \frac{1}{\sigma \sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2}\right)$$

Regression

$$y = \beta_0 + \beta_1 x + \epsilon$$

Further Work

1. Regression with multiple variables
2. Linear Regression with Ridge and Lasso
3. Logistic Regression and the Softmax Function
4. Support Vector Machines
5. Decision Trees and Random Forests
6. Boosting and Gradient Descent

Admissible Regions





Claude Shannon - "Father of information theory".

Decoding Data Quality:

An introduction to Entropy and Mean Square Error (MSE).

A Mathematical
Theory of
Communication By C.
E. SHANNON

- Define what the Mutual Information is.
- Interpretation of the MSE can be interpreted as the expected variance of the estimation error.
- Mutual Information can be adequate for feature selection only when the conditional distribution of the estimation error is Gaussian,

What is Mutual Information?

Mutual information (Shannon, 1948) measures the dependency between two random variables.

Let X and Y be random variables with probability density functions f_X and f_Y , and domains \mathcal{X} and \mathcal{Y} . The mutual information between X and Y is defined as:

$$I(X; Y) = - \int_{\mathcal{X}} \int_{\mathcal{Y}} f_{X,Y}(x, y) \log \frac{f_{X,Y}(x, y)}{f_X(x)f_Y(y)} dx dy.$$

Which can be written as:

$$I(X; Y) = H(Y) - H(Y|X).$$

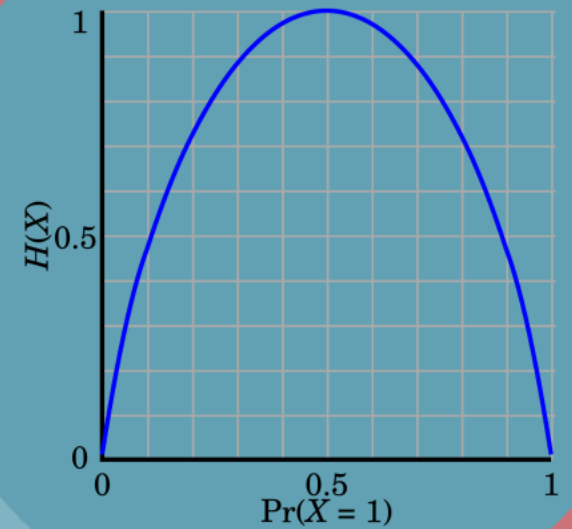
$$H(X) = - \int_{\mathcal{X}} f_X(x) \log f_X(x) dx$$

and

$$H(Y|X) = - \int_{\mathcal{X}} \int_{\mathcal{Y}} f_{X,Y}(x, y) \log \frac{f_X(x)}{f_{X,Y}(x, y)} dx dy.$$

What is Entropy?

In simple terms, entropy measures the uncertainty in a set of possible outcomes.



Entropy in a Bernoulli trial X (a random experiment in which X can take the values 0 or 1). The entropy depends on the probability $P(X=1)$ that X takes the value 1. When $P(X=1)=0.5$, all possible outcomes are equally likely, so the result is unpredictable and the entropy is at its maximum.

Regression

Let us assume that for a given subset of d features, the output $Y \in \Re$ depends probabilistically on the input $X \in \Re^d$. Moreover, the function f provides an estimate $\hat{Y} = f(X)$ of Y given X . Then, the estimation error is

$$\epsilon = f(X) - Y$$

For a given estimate of f , and with some algebra it can be seen that the conditional entropy of Y given x can be written as:

$$\begin{aligned} H(Y|X) &= \int_{\mathcal{X}} f_X(x) H(Y|X=x) dx \\ &= \int_{\mathcal{X}} f_X(x) H(f(X) + \epsilon | X=x) dx. \end{aligned}$$

The Normal Case

1.1. The normal Case

So far, we only discussed the case where the random variables are discrete. However, in regression problems, a more common scenario is to assume normality in our error ϵ . It is convenient to set a baseline for our measurement of our MSE by assuming normality in the distribution of our errors. Let $\epsilon \sim N(0, \sigma^2)$, we write

$$\epsilon \sim \phi(\epsilon) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{\epsilon^2}{2\sigma^2} \right\}.$$

Where the distribution of $\epsilon \sim N(0, \sigma^2)$, we can express $H(Y | X)$ as

$$\begin{aligned} H(\epsilon | X) &= - \int \phi(\epsilon) \ln \phi(\epsilon) d\epsilon \\ &= - \int \phi(\epsilon) \left[-\frac{\epsilon^2}{2\sigma^2} - \ln \sqrt{2\pi\sigma^2} \right] \\ &= \frac{1}{2} + \frac{1}{2} \ln(2\pi\sigma^2) \\ &= \frac{1}{2} \ln(2\pi e\sigma^2) \end{aligned} \tag{6}$$

The MSE can be interpreted as the expected variance of the estimation error. Since the estimation error is assumed to be identically distributed for any $x \in \mathcal{X}$, its variance is precisely equal to the MSE. For the normal case, the variance can be written in terms of σ^2 solely.

Admissible Regions

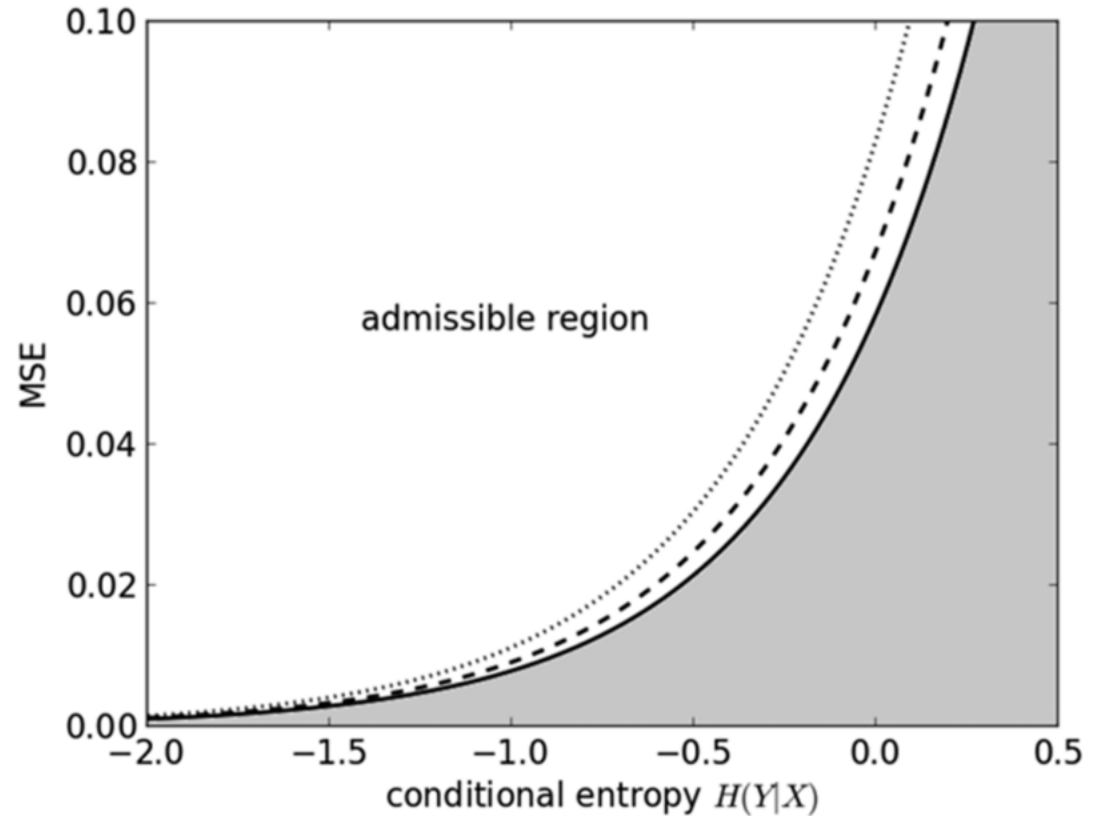
Since the Gaussian distribution is the maximum entropy distribution for a given estimation error variance σ^2 (?), the curve corresponding to the Gaussian error gives a lower bound for the MSE and defines an admissible region for the MSE.

Using the expressions for the conditional target entropies and these relationships it is possible to express the MSE in sole terms of $H(Y|X)$. For the normal case, the MSE becomes

$$MSE = \frac{1}{2\pi e} \exp \{2H(Y|X)\}$$

and similar the Root Mean Square error (RMSE) as

$$RMSE = \sqrt{MSE}.$$





(Short) Discussion

Mutual information is an important criterion for feature selection, equivalent to minimizing the conditional target entropy $H(Y|X)$, which corresponds to minimizing the MSE.

Further Work

Children's Work and Child
Labour: Prevalence Rates and
The Importance of Plural
Causality work with Professor
Wendy Olsen

Child labor complexity: Olsen's research shows the need for a multifaceted, integrated strategy to address child labor.

Economic, sociological, and policy perspectives provide a comprehensive understanding for creating sustainable solutions.



Q&A

Thanks ! Any questions?



Social Statistics - Away Day

Dr Diego Perez Ruiz

May / 2024

Further Work

Read the book "The Elements of Statistical Learning" by Tibshirani, Hastad, and Tibshirani. It is a great resource for understanding the theoretical foundations of machine learning.

(Short) Discussion

Mutual information is an important criterion for feature selection, equivalent to minimising the conditional target entropy $H(Y|X)$, which corresponds to minimising the MSE.

Q&A

Thanks! Any questions?



Decoding Data Quality: An Introduction to Entropy and Mean Square Error (MSE).

Entropy and the Mean Square Error (MSE) are two fundamental concepts in the field of data science. Entropy measures the uncertainty or information content of a random variable, while MSE measures the average squared difference between the estimated values and the true values.

A Mathematical Theory of Communication By C. E. SHANNON

What is Mutual Information?

Mutual information (MI) is a measure of the dependence between two random variables. It quantifies the amount of information that one random variable contains about another. MI is symmetric, meaning $I(X;Y) = I(Y;X)$.

What is Entropy?

In simple terms, entropy measures the uncertainty or randomness in a set of possible outcomes.



Entropy is a measure of the average amount of information contained in a random variable. It is calculated as the negative logarithm of the probability of each outcome, weighted by the probability of that outcome.

Regression

Regression is a statistical method used to model the relationship between a dependent variable and one or more independent variables. It is commonly used for predicting the value of the dependent variable based on the values of the independent variables.

The Normal Case

The normal distribution is a probability distribution that is symmetric and bell-shaped. It is characterized by its mean and standard deviation. The normal distribution is a special case of the more general Gaussian distribution.

Adaptive Regression

