Social Statistics - Away Day

Dr Diego Perez Ruiz

May / 2024

Decoding Data Quality:

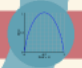
An introduction to Entropy and Mean Square Error (MSE).

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 adequate for feature selection only when the conditional distribution of the estimation error is Gaussian.

What is Mutual Information?

What is Entropy?



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

Regression

The Normal Case

Admissible Regions

Further Work

(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.

Q&A

Thanks ! Any questions?

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

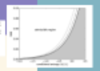
- Define what the Mutual Information is.
- Interpretation of the MSE can be interpreted as the expected variance of the estimation error.
- Mutual information can be adequate for feature selection only when the conditional distribution of the estimation error is Gaussian.

What is Mutual Information?

Regression

The Normal Case

Admissible Regions



Further Work

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

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:

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

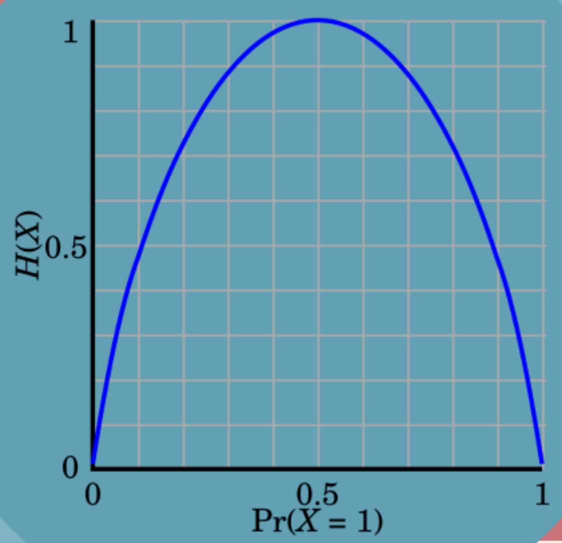
Which can be written as: $I(X;Y) = H(Y) - H(Y|X).$

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

and

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

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 $H(Y \mid X)$ as

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

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

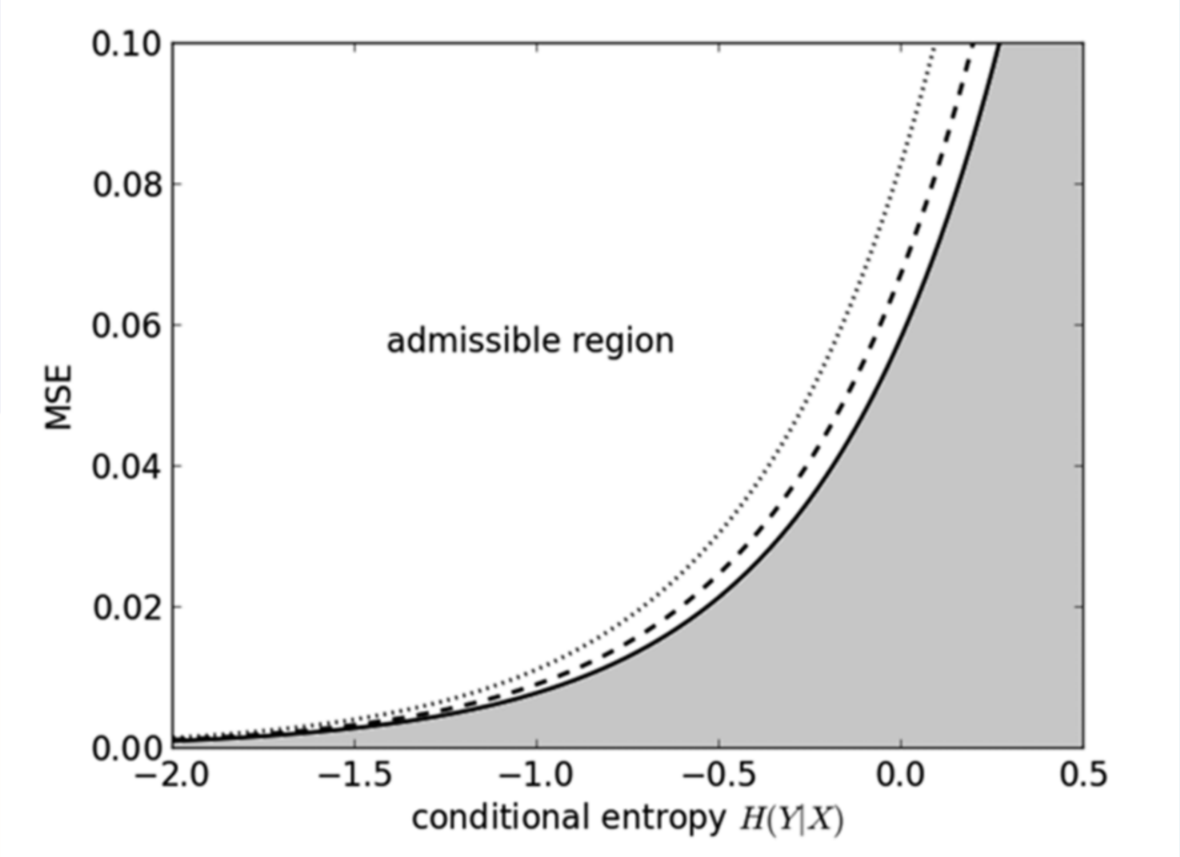
Since the Gaussian distribution is the maximum entropy distribution for a given estimation error variance σ^2 (?), the curve corresponding to the Gaussian error gives a lower bound for the MSE and defines an admissible region for the MSE.

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

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

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

Decoding Data Quality:

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

A Mathematical Theory of Communication By C. E. SHANNON

What is Mutual Information?

What is Entropy?

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

The Normal Case

Regression

Admissible Regions

Further Work

(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.

Q&A

Thanks! Any questions?