25. Welches Ziel hat die Regressionsanalyse? Gehen Sie sowohl auf die Regression im Allgemeinen als auch auf die lineare Regression ein. Was passiert bei der Berechnung linearer Regression wenn bei den Messdaten exakt 1 Messfehler auftritt und statt einem durchschnittlichen Wert ein extrem hoher Wert gemessen wird (Beispielsweise statt 15 kommt 15.000.000)?

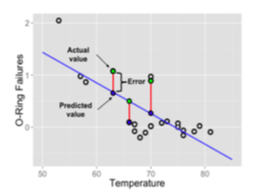
Regressionsanalysen sind statische Analysen. Sie stellen die Beziehung zwischen einer abhängigen und einer oder mehrerer unabhängigen Variablen her.

Eine lineare Regression definiert die Abhängigkeit einer Variable Y zu einer unabhängigen Variable X. Wären es mehrere unabhängige Variablen X, dann sprechen wir von einer multiplen linearen Regression.

Erwische ich genau diesen Messwert, kann ich das mit der linearen Regression nicht ausbessern. Das Messergebnis ist nicht exakt. Habe ja nur eine unabhängige Variable. Hätte ich mehrere unabhängige Variablen wie bei der multiplen linearen Regression könnte ich diesen einen Messwert vernachlässigen. - ????

26. Was ist Ziel des OLS (Ordinary least Square; de: Methode der kleinsten Quadrate) – Vorgehens? Erklären Sie dabei auch den Begriff Residuum.

Es wird versucht die Abweichung der Gerade zwischen den Messpunkten zu minimieren. Die Gerade soll so nah wie möglich an die Messpunkte angeglichen werden. Residuum ist dabei der Abstand/Abweichung von der geschätzten Gerade zum tatsächlichen Messpunkt.



27. Was besagt der „correlation coefficient“? Erklären Sie dabei   
a) welchen Sinn der Korrelationskoeffizient hat,   
b) welche Einheit,   
c) welchen Wertebereich,   
d) was die Extremwerte und 0 bedeuten,   
e) unter welchen Randbedingungen er Anwendung findet und   
f) ob der Korrelationskoeffizient zwischen X und Y   
 gleich dem Korrelationskoeffizienten zwischen Y und X ( rxy ?= ryx).

Der „correlation coefficient“ erklärt den Zusammenhang zwischen zwei intervallskalierten Merkmalen.

1. Er bestimmt in wie weit ein linearer Zusammenhang zu einem Messwert besteht. Im Fall der linearen Regression kann damit bestimmt werden, ob der Messwert wichtig für die Voraussage ist oder nicht.
2. Ein Maß für den Grad des linearen Zusammenhangs???
3. Von -1 bis +1
4. Extremwerte: vollständiger linearer Zusammenhang (bei +1 steigt die Gerade an, bei -1 sinkt die Gerade). 0: beide Merkmale hängen gar nicht linear voneinander ab.
5. Es muss eine lineare Abhängigkeit gegeben sein???
6. Ja es ist gleich

28. Was ist „Entropy“? Erklären Sie in Folge die Entropie anhand des konkreten Beispiels „Würfeln mit Spezialwürfel“. Der Spezialwürfel hat dabei 2 blaue Seiten und 4 rote. Erklären Sie „Information gain“.

Die Entropie beschreibt die Menge an Zufall, die in einem oder mehreren zufälligen Ergebnissen steckt. Die Entropie wäre 0 wenn ein sicheres Ergebnis oder eine Folge von sicheren Ergebnissen vorliegt.

Würfel = 6 Seiten

Blau = 2/6

Rot = 4/6

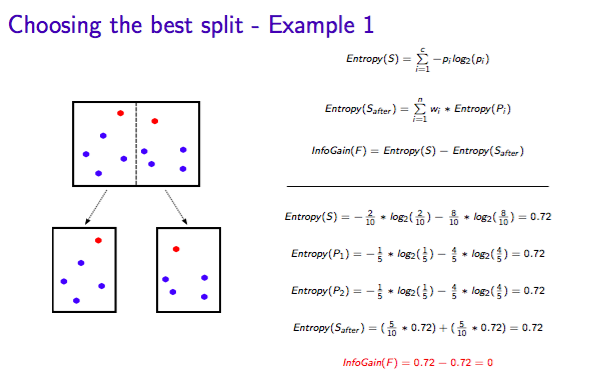
Entropie = - (p \* log2 (p) + q \* log2 (q)) =

- (4/6 \* log2 (4/6) + 2/6 \* log2 (2/6)) =

- (4/6 \*( -0.584963) + 2/6 \*( -1.584963)) =

-(0.0817036666666 - 1,2516296666666) = 1.169926

Information Gain: Misst eine Veränderung in der Homogenität einer Teilung von Daten segmenten. ???



Das wäre der ideale Split, damit wäre Info gain 0. Im Beispiel mit den Würfeln würde jedes Segment jeweils eine blaue Seite beinhalten und 2 rote.