# Towards interpretable machine-learning for materials science

Coffee Talk

Benjamin Regler
Fritz Haber Institute of the Max Planck Society, Berlin, Germany

June 3, 2019
**(Berlin)**

MAX-PLANCK-GESELLSCHAFT

**Definition:**
Interpretable machine-learning focuses on the extraction of relevant knowledge about domain relationships contained in data.

**Definition:**
Interpretable machine-learning focuses on the extraction of relevant knowledge about domain relationships contained in data.

**Find functional relationships in materials science data**

1. Quantify knowledge (=information) that is shared between features and target material property (quantification of relationships)
2. Determine strength of relationships (ranking)

**Def.: Information**

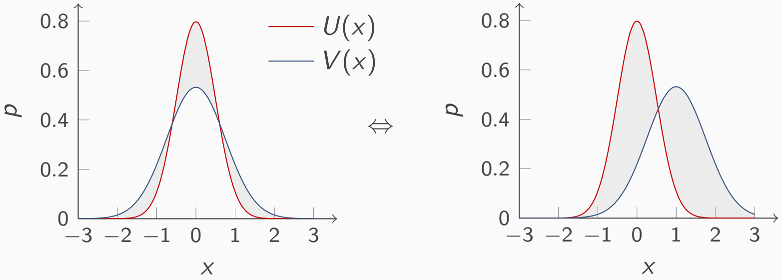
Quantity that encodes the difference between distributions $U$ and $V$ via Kullback-Leibler divergence $D_{\mathrm{KL}}(U \,\|\, V)$

$\sim$ disagreement of distributions

**Def.: Information**

Quantity that encodes the difference between distributions $U$ and $V$ via Kullback-Leibler divergence $D_{\mathrm{KL}}(U \| V)$

$\sim$ disagreement of distributions




low $D_{KL}$ high $D_{KL}$

FRITZ-HABER-INSTITUT
MAX-PLANCK-GESELLSCHAFT

**Def.: Information**

Quantity that encodes the difference between distributions $U$ and $V$ via Kullback-Leibler divergence $D_{\mathrm{KL}}(U\,\|\,V)$

$\sim$ disagreement of distributions

**(Cumulative) mutual information**

Quantifies information that is shared between two variables $X$ and $Y$:

$P(x) := \int_{-\infty}^{x} p(x')\,dx'$

$$D_{\mathrm{KL}}(P(y,x)||P(y)P(x)) \equiv \mathcal{I}(Y;X)$$

**Def.: Information**

Quantity that encodes the difference between distributions $U$ and $V$ via Kullback-Leibler divergence $D_{\mathrm{KL}}(U \| V)$

$\sim$ disagreement of distributions

**(Cumulative) mutual information**

Quantifies information that is shared between two variables $X$ and $Y$:

$P(x) := \int_{-\infty}^{x} p(x')\, dx'$

$$D_{\mathrm{KL}}(P(y,x)||P(y)P(x)) \equiv \mathcal{I}(Y;X)$$

Cumulative entropy

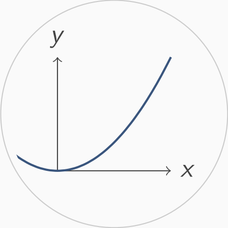
$$\mathcal{I}(Y;X) = \int_{y\in Y}\int_{x\in X} P(y,x) \log \frac{P(y,x)}{P(y)P(x)}\, dx\, dy = \mathcal{H}(Y) - \mathcal{H}(Y|X)$$

Conditional cumulative entropy

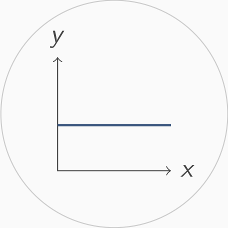
# Information theory



Target ($Y$)



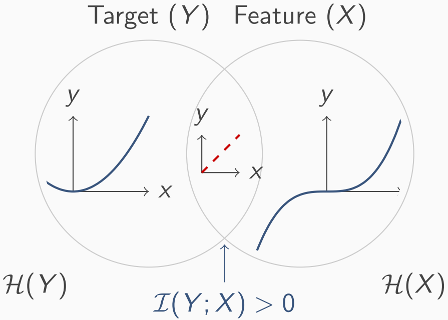

Feature ($X$)

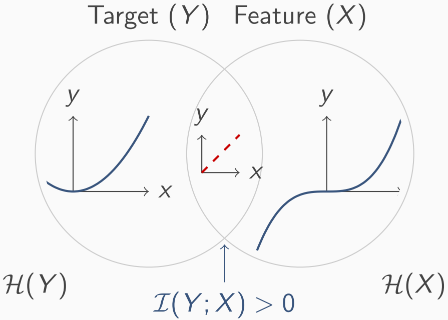



$\mathcal{H}(Y)$

$\mathcal{I}(Y;X) = 0$

$\mathcal{H}(X)$


Target (Y)
Feature (X)
y
x
y
x
y
H(Y)
I(Y; X) > 0
H(X)

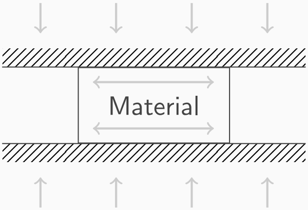



Compute dependence measure $\mathcal{D}(Y;X) := \frac{\mathcal{I}(Y;X)}{\mathcal{H}(Y)}$ for <u>every</u> combination of feature subsets $\mathbf{X} = \{X_1, \ldots, X_d\}$

$\rightarrow$ Select subset $\mathbf{X}^*$ with the highest contribution to $Y$, i.e., features which contribute most to the material's property prediction.

# Compressive strength of concrete[1]

» Resistance of a material to breaking under compression




- 1030 samples, 7+1 features (components+age)
- **Components:** cement, fly ash, blast furnace slag, water, superplasticizer, coarse and fine aggregate

**Task:** Find optimal mix proportion of ingredients to predict compressive strength of concrete.

[1] UCI Machine Learning Repository: `http://archive.ics.uci.edu/ml/datasets/concrete+compressive+strength`

FRITZ-HABER-INSTITUT
MAX-PLANCK-GESELLSCHAFT

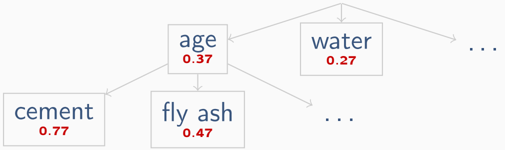
## Subspace search


[1] UCI Machine Learning Repository: http://archive.ics.uci.edu/ml/datasets/concrete+compressive+strength

# Compressive strength of concrete[1]

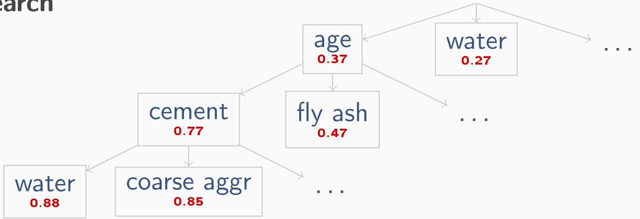


## Subspace search




[1] UCI Machine Learning Repository: `http://archive.ics.uci.edu/ml/datasets/concrete+compressive+strength`

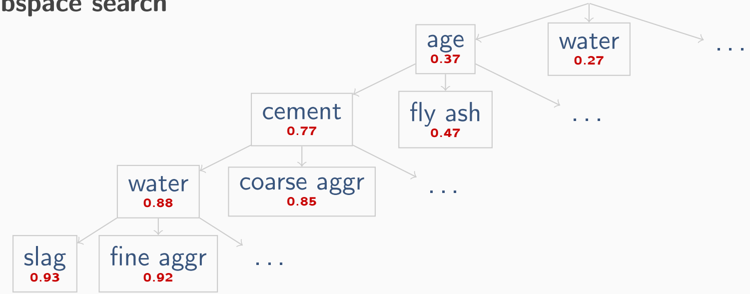
## Subspace search




[1] UCI Machine Learning Repository: http://archive.ics.uci.edu/ml/datasets/concrete+compressive+strength

FRITZ-HABER-INSTITUT
MAX-PLANCK-GESELLSCHAFT

## Subspace search

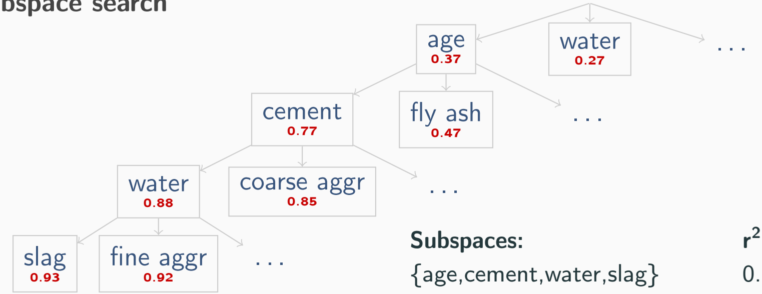



[1] UCI Machine Learning Repository: `http://archive.ics.uci.edu/ml/datasets/concrete+compressive+strength`

# Compressive strength of concrete[1]



## Subspace search



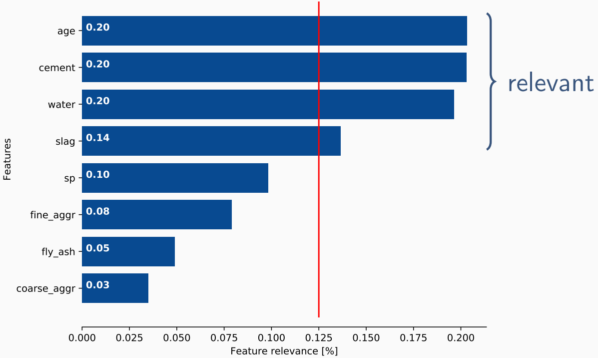

| **Subspaces:** | $\mathbf{r^2}$ |
|---|---|
| {age,cement,water,slag} | 0.93 |
| {age,cement,water,fine aggr} | 0.92 |
| ... | ... |

→ Subspace search finds non-unique sparse representations

[1] UCI Machine Learning Repository: `http://archive.ics.uci.edu/ml/datasets/concrete+compressive+strength`

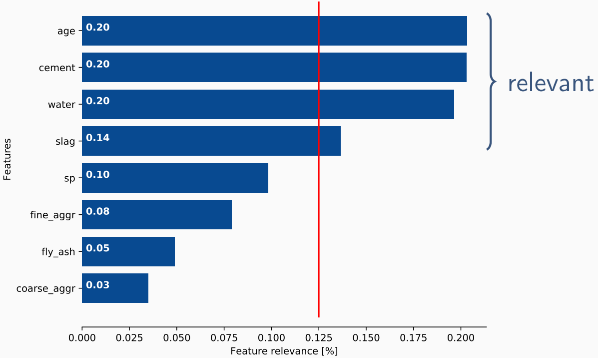
# Compressive strength of concrete[1]

Transform tree representation of subspaces into feature relevance histogram:




[1] UCI Machine Learning Repository: `http://archive.ics.uci.edu/ml/datasets/concrete+compressive+strength`

Transform tree representation of subspaces into feature relevance histogram:




| | $r^2$ | RMSE [GPa] | MAE [GPa] | MaxAE [GPa] |
|---|---|---|---|---|
| Full model [8 features] | $0.94 \pm 0.02$ | $3.98 \pm 0.60$ | $2.60 \pm 0.28$ | $18.36 \pm 6.62$ |
| Reduced model [age,cement,water,slag] | $0.93 \pm 0.02$ | $4.40 \pm 0.65$ | $2.94 \pm 0.32$ | $19.63 \pm 7.61$ |

[1] UCI Machine Learning Repository: http://archive.ics.uci.edu/ml/datasets/concrete+compressive+strength

# Machine-learning models