Machine Learning in Medical Applications: from fundamental topics to practical aspects

Kelwin Fernandes

INESC TEC

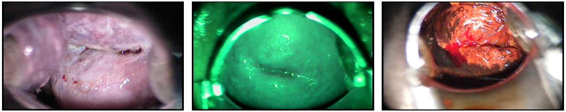
kelwinfc@gmail.com

June 12, 2018

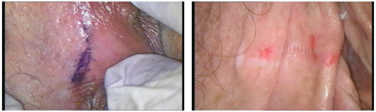
The following pictures may be disturbing to some audiences.
Viewer discretion is advised

Digital Colposcopy

- **Cervical cancer screening** (Dr. J. Fernandes, HUC, Venezuela)



- **Forensic assessment of sexual assault** (Dr. B.S. Astrup, SDU, Denmark).



Relevant Properties:

- Partially observable data
- Imbalance distribution
- Scarce data
- Sensitive data
- Catastrophic failure

Similar contexts:

- Credit Fraud
- Intruder detection
- Churn

The standard Data Science toolkit

- Supervised learning
  - Classification: binary, multiclass
  - Regression
- Unsupervised
  - Clustering/Compression
- Reinforcement Learning

The standard Data Science toolkit

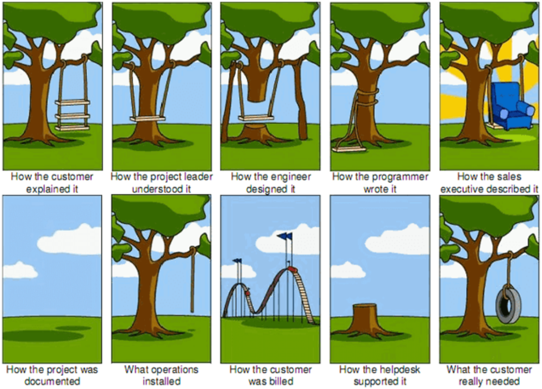
- Supervised learning
  - Classification: binary, multiclass
  - Regression
- Unsupervised
  - Clustering/Compression
- ~~Reinforcement Learning~~

The standard Data Science toolkit

- Supervised learning
  - Classification: binary, multiclass
  - Regression
- Unsupervised
  - Clustering/Compression
- ~~Reinforcement Learning~~

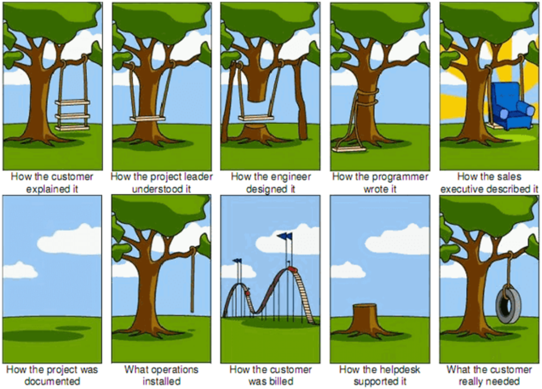
```
import sklearn
import tensorflow as tf
import keras
...
model.fit(X, y)
pred = model.predict(Xtest)
if pred == 1:
   print(Take some benuron!)
```

The standard Data Science product



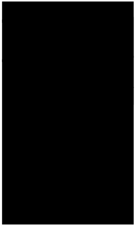
The more decisions you make,
the more you distance your model from reality

The standard Data Science product

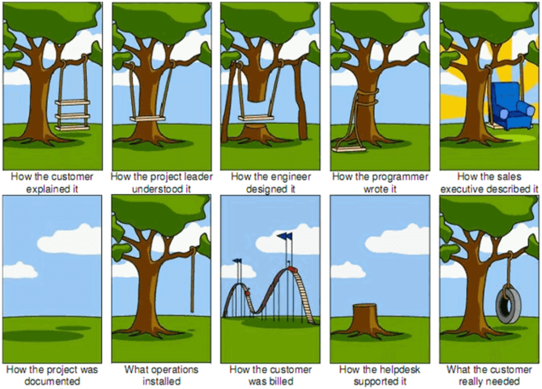


The more decisions you make,
the more you distance your model from reality

Your ML-based solution



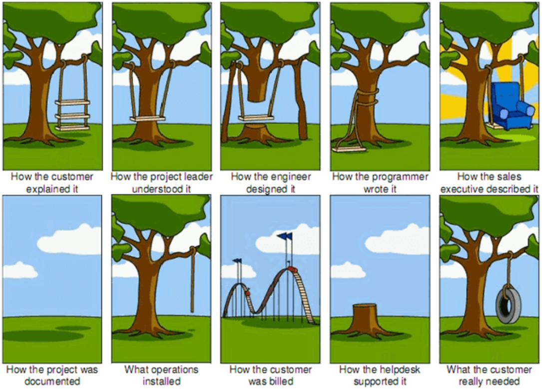
The standard Data Science product



The more decisions you make,
the more you distance your model from reality

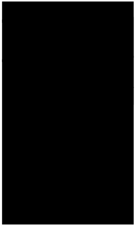
Your ML-based solution

The standard Data Science product

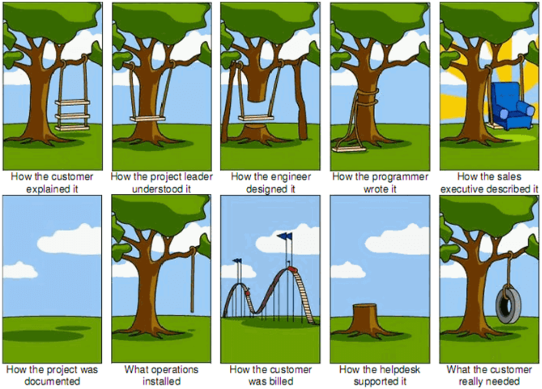


The more decisions you make,
the more you distance your model from reality

Your ML-based solution

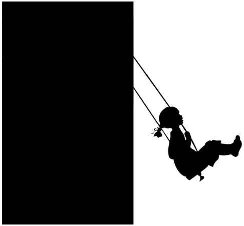


The standard Data Science product

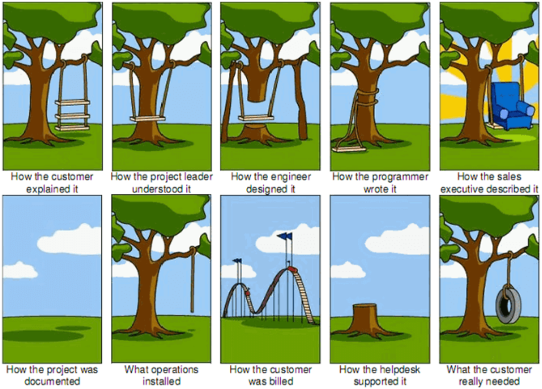


The more decisions you make,
the more you distance your model from reality

Your ML-based solution

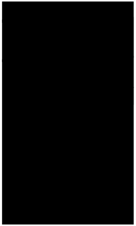


The standard Data Science product

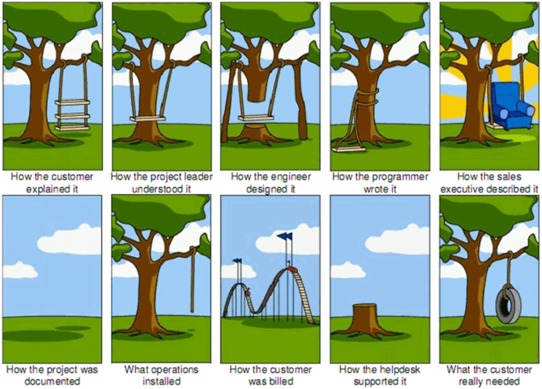


The more decisions you make,
the more you distance your model from reality

Your ML-based solution



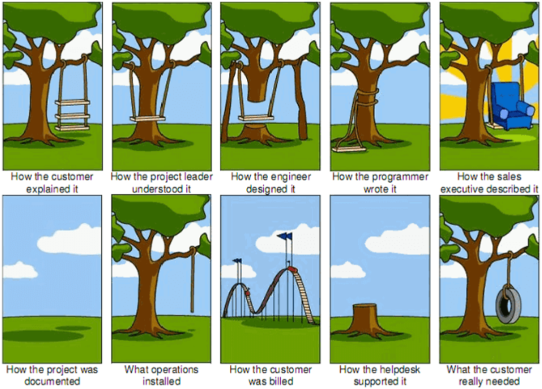
The standard Data Science product



The more decisions you make,
the more you distance your model from reality

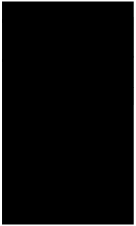
Your ML-based solution

The standard Data Science product

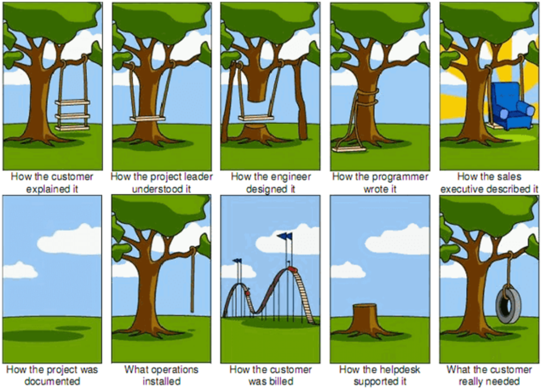


The more decisions you make,
the more you distance your model from reality

Your ML-based solution

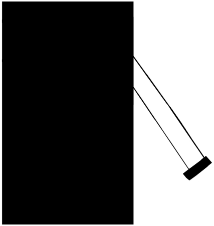


The standard Data Science product

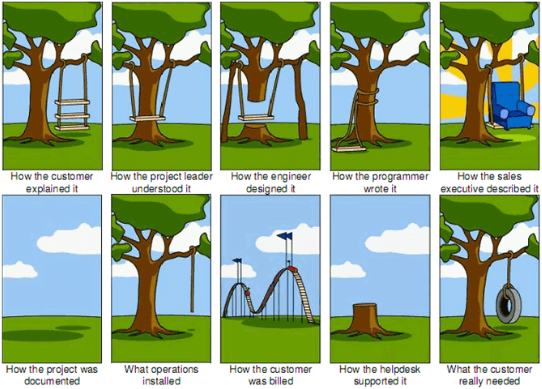


The more decisions you make,
the more you distance your model from reality

Your ML-based solution

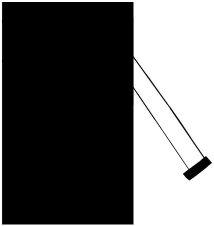


The standard Data Science product



The more decisions you make,
the more you distance your model from reality

Your ML-based solution



CAD system for cancer recognition

Task

Predict if a patient has cancer.

Learning strategy

???

.

.

CAD system for cancer recognition

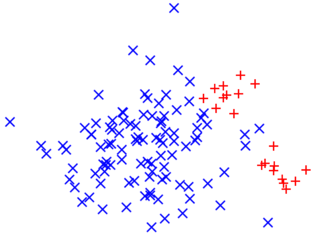
Task

Predict if a patient has cancer.

Learning strategy

Binary classification (Pos/Neg)

Positive=Cancer vs.
Negative=Healthy



Your dataset today

CAD system for cancer recognition

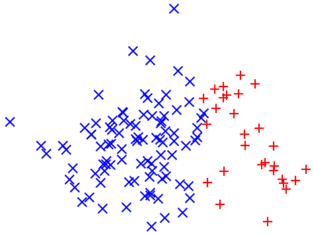
Task

Predict if a patient has cancer.

Learning strategy

Binary classification (Pos/Neg)

Positive=Cancer vs.
Negative=Healthy



Your dataset
K years later

CAD system for cancer recognition

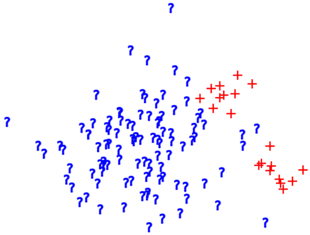
Task

Predict if a patient has cancer.

Learning strategy

Positive-Unlabeled classification

Cancer vs. Rest
(binary class + prob calibration)

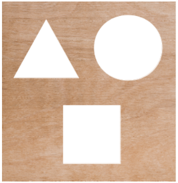


Have cancer vs.
don't have cancer yet

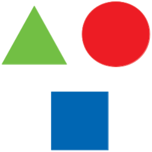
Elkan, Charles, and Keith Noto. "Learning classifiers from only positive and unlabeled data." Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2008.

What happened?

standard ML toolkit



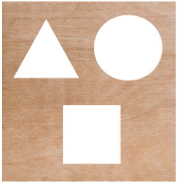
standard ML problems



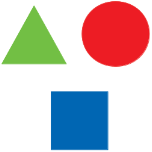
a wild problem appears

What happened?

standard ML toolkit



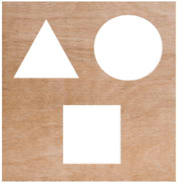
standard ML problems



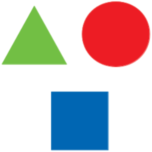
a wild problem appears

What happened?

standard ML toolkit



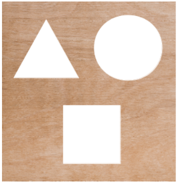
standard ML problems



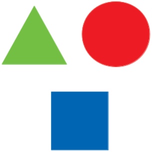
a wild problem appears

What happened?

standard ML toolkit



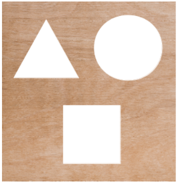
standard ML problems



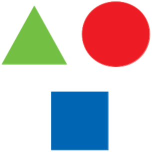
a wild problem appears

A better approach...

standard ML toolkit



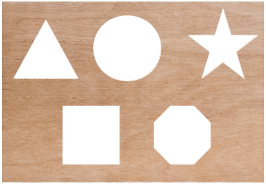
standard ML problems



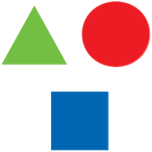
a wild problem appears

A better approach...

extended ML toolkit



standard ML problems



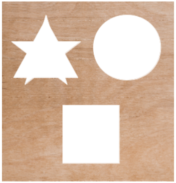
a wild problem appears

Read the literature

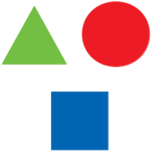
Binary, Multiclass, . . . Positive-Unlabeled, Multilabel, Ordinal, Multiple-Instance, Ranking

A better approach...

adapted ML toolkit



standard ML problems



a wild problem appears

Build new models

CAD system for cancer recognition

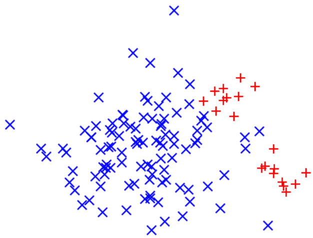
Task

Predict if a patient has cancer.

Learning strategy

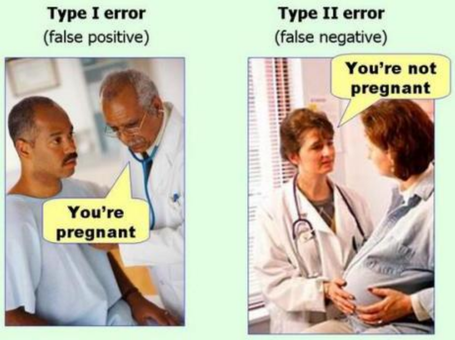
Binary classification (Pos/Neg)

Positive=Cancer vs.
Negative=Healthy

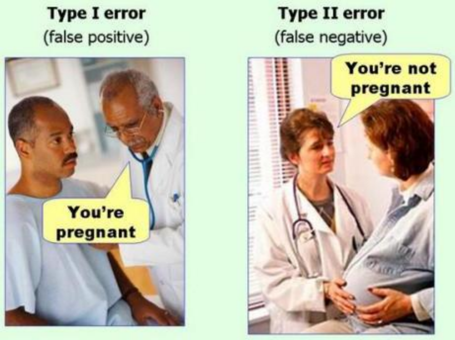


Your dataset today

CAD system for cancer recognition... filtering



CAD system for cancer recognition... filtering



Constraining Type II Error

Maximize TNR
subject to FNR $\leq \rho$.

- Filter easy cases that don't have the disease.
- builds trust.

Cruz, Ricardo, Kelwin Fernandes, Joaquim F. Pinto Costa, and Jaime S. Cardoso. "Constraining Type II Error: Building Intentionally Biased Classifiers." In International Work-Conference on Artificial Neural Networks, pp. 549-560. Springer, Cham, 2017.

Imbalanced classification

- **Screening program for breast/cervical/<you name it> cancer:**
  - Healthy patients: 99.9%
  - Sick patients: 0.1%

- **Approaches**:

Imbalanced classification

- **Screening program for breast/cervical/<you name it> cancer:**
  - Healthy patients: 99.9%
  - Sick patients: 0.1%

- **Approaches**:
  - Resampling (Undersampling/Oversampling):

Imbalanced classification

- **Screening program for breast/cervical/<you name it> cancer:**
  - Healthy patients: 99.9%
  - Sick patients: 0.1%

- **Approaches**:
  - Resampling (Undersampling/Oversampling):
    - Artificial problem.
    - Doesn't resemble the population distribution.

Imbalanced classification

- **Screening program for breast/cervical/<you name it> cancer:**
  - Healthy patients: 99.9%
  - Sick patients: 0.1%

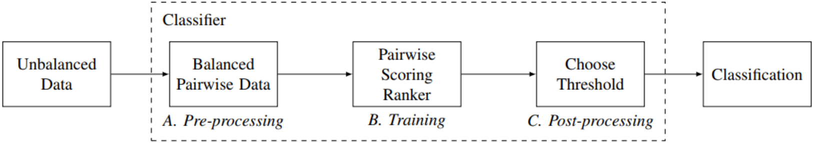
- **Approaches**:
  - Resampling (Undersampling/Oversampling):
    - Artificial problem.
    - Doesn't resemble the population distribution.
  - Cost matrix
    - What's the cost of killing a patient?
    - Hard acceptance by medical teams.
    - . . . Not so much in SNS.

Imbalanced classification

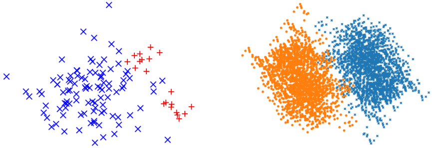
- **Screening program for breast/cervical/<you name it> cancer:**
  - Healthy patients: 99.9%
  - Sick patients: 0.1%

- **Approaches**:
  - Resampling (Undersampling/Oversampling):
    - Artificial problem.
    - Doesn't resemble the population distribution.
  - Cost matrix
    - What's the cost of killing a patient?
    - Hard acceptance by medical teams.
    - ... Not so much in SNS.
  - One-class/Outlier
    - Data is new gold.
    - You don't want to throw it away.

Imbalanced classification as Ranking

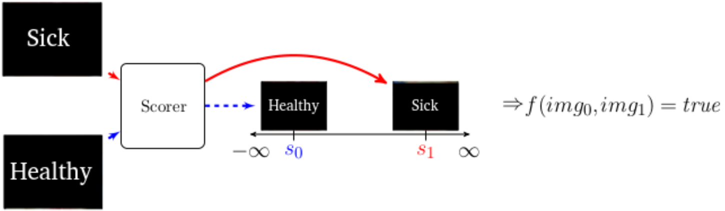


- Pairwise scoring ranker: $f(a, b) = s(a) < s(b)$
- Balanced by definition $f(a, b) = 0, f(b, a) = 1$
- Linear model: $\omega \cdot (x_0 - x_1) > 0 \equiv \omega \cdot x_0 > \omega \cdot x_1 \equiv s(x_0) > s(x_1)$

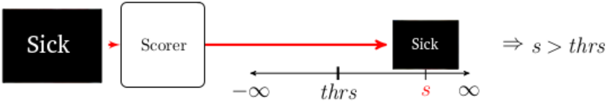


Cruz, Ricardo, Kelwin Fernandes, Jaime S. Cardoso, and Joaquim F. Pinto Costa. "Tackling class imbalance with ranking." In Neural Networks (IJCNN), 2016 International Joint Conference on, pp. 2182-2187. IEEE, 2016.

Imbalanced classification as Ranking

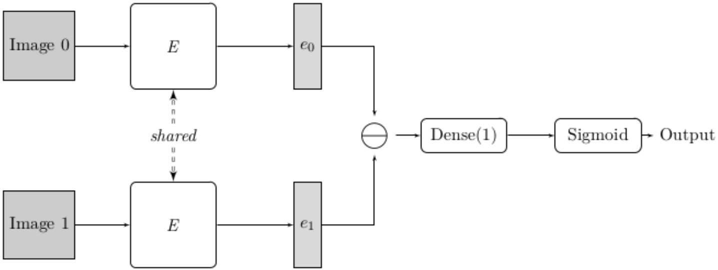


(a) Training pairwise scoring rankers.



(b) Inference with pointwise scorer obtained from pairwise scoring rankers.

Deep Ranking Network



Fernandes, Kelwin, Jaime S. Cardoso, and Birgitte Schmidt Astrup. "A deep learning approach for the forensic evaluation of sexual assault." Pattern Analysis and Applications (2018): 1-12.

Medical data is scarce

- Use simple models: linear models, decision trees, etc.
- Use complex models + regularization.
- **Transfer Learning**/Multitask learning
  - Can be understood as regularization

Medical data is scarce... Transfer learning

```
model = ResNet50(weights='imagenet')
model.predict(PatientData)
```

Medical data is scarce... Transfer learning

```
model = ResNet50(weights='imagenet')
model.predict(PatientData)
```

Medical data is scarce... Transfer learning

```
model = ResNet50(weights='imagenet')
model.predict(PatientData)
```

 ... 

Medical data is scarce... Transfer learning

```
base= ResNet50(weights='imagenet',include_top=False)
out = Dense(1, activation='sigmoid')(base.output)
model = Model(inputs=base.input, outputs=out)

model.fit(trX, trY)
model.predict(PatientData)
```

Medical data is scarce... Transfer learning

```
base= ResNet50(weights='imagenet',include_top=False)
out = Dense(1, activation='sigmoid')(base.output)
model = Model(inputs=base.input, outputs=out)

model.fit(trX, trY)
model.predict(PatientData)
```

Medical data is scarce... Transfer learning

```
base= ResNet50(weights='imagenet',include_top=False)
out = Dense(1, activation='sigmoid')(base.output)
model = Model(inputs=base.input, outputs=out)

model.fit(trX, trY)
model.predict(PatientData)
```

...

Medical data is scarce... Transfer learning

```
base= ResNet50(weights='imagenet',include_top=False)
out = Dense(1, activation='sigmoid')(base.output)
model = Model(inputs=base.input, outputs=out)

for layer in base.layers:
   layer.trainable = False

model.fit(trX, trY)
model.predict(PatientData)
```

Medical data is scarce... Transfer learning

```
base= ResNet50(weights='imagenet',include_top=False)
out = Dense(1, activation='sigmoid')(base.output)
model = Model(inputs=base.input, outputs=out)

for layer in base.layers:
   layer.trainable = False

model.fit(trX, trY)
model.predict(PatientData)
```

Medical data is scarce... Transfer learning

```
base= ResNet50(weights='imagenet',include_top=False)
out = Dense(1, activation='sigmoid')(base.output)
model = Model(inputs=base.input, outputs=out)

for layer in base.layers:
   layer.trainable = False

model.fit(trX, trY)
model.predict(PatientData)
```

 ... 

Medical data is scarce... Transfer learning

- Transfer high-level properties instead of initialization values.
- Regularization:

$$M^* = \text{argmax}_M \left(P(M|D^{target}) + \lambda \text{similarity}(M, M^{source}) \right)$$

- **Advantages:**
  - Trade-off between target data and source priors.
  - Wide range of "similar" applications as you increase the property abstraction.
  - Security/Privacy: You only reveal the properties you want to share.

Fernandes, Kelwin, and Jaime S. Cardoso. "Hypothesis transfer learning based on structural model similarity." Neural Computing and Applications: 1-14.

Medical data is scarce... Transfer learning

- **Value:**

$$\sum_{i=1}^{d}(\omega_i^t - \omega_i^s)^2$$

  - **Source:** Smoking and lung cancer $(+0.6) \implies$ **Target:** cervical (~ 0.6)

Medical data is scarce... Transfer learning

- **Value:**

$$\sum_{i=1}^{d}(\omega_i^t - \omega_i^s)^2$$

  - **Source:** Smoking and lung cancer $(+0.6)$ $\implies$ **Target:** cervical (~ 0.6)

- **Sign:**

$$\sum_{i=1}^{d}\max(0, -\omega_i^t \cdot \mathrm{sign}(\omega_i^s))$$

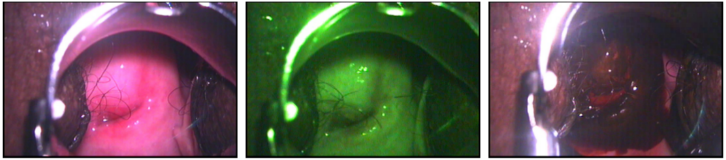
  - **Source:** Smoking and lung cancer $(+0.6)$ $\implies$ **Target:** breast (positive)

Fernandes, Kelwin, Jaime S. Cardoso, and Jessica Fernandes. "Transfer learning with partial observability applied to cervical cancer screening." In Iberian conference on pattern recognition and image analysis, pp. 243-250. Springer, Cham, 2017.

Medical data is scarce... Transfer learning

- **Quality Assessment of Digital Colposcopy**
  - Features: occlusion, blurriness, artifacts, etc.

  - Subjective concept
  - **Value:** changes from expert to expert.
  - **Sign:** holds between experts.

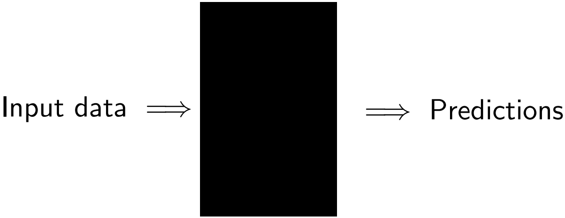


- **Sign transfer:**
  - Inter-expert
  - Inter-modality

Fernandes, Kelwin, Jaime S. Cardoso, and Jessica Fernandes. "Transfer learning with partial observability applied to cervical cancer screening." In Iberian conference on pattern recognition and image analysis, pp. 243-250. Springer, Cham, 2017.

Predictions aren't enough... tell my why!

- Trust is crucial when dealing with medical applications (and doctors).
- Try to open the black box



- Identify predictions that are incorrect.
- Design a treatment plan.
- Identify patterns that aren't known by the doctors.

Predictions aren't enough... tell my why!

- Trust is crucial when dealing with medical applications (and doctors).
- Try to open the black box

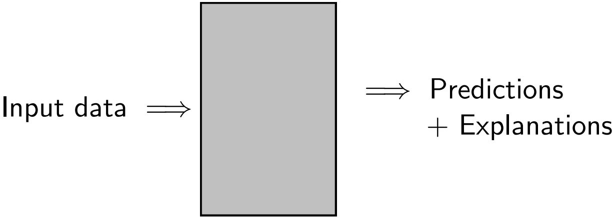
Input data $\Longrightarrow$ $\boxed{}$ $\Longrightarrow$ Predictions

$+$ Explanations

- Identify predictions that are incorrect.
- Design a treatment plan.
- Identify patterns that aren't known by the doctors.

Predictions aren't enough... tell my why!

- Trust is crucial when dealing with medical applications (and doctors).
- Try to open the black box



- Identify predictions that are incorrect.
- Design a treatment plan.
- Identify patterns that aren't known by the doctors.

Predictions aren't enough... tell my why!

- **Interpretable models:**
  - Linear models (e.g., linear regression, logistic regression).
  - Decision Trees.
  - Scorecards.
  - Inductive rules.
  - Case-based (e.g. KNN)

Predictions aren't enough... tell my why!

- **Interpretable models:**
  - Linear models (e.g., linear regression, logistic regression).
  - Decision Trees.
  - Scorecards.
  - Inductive rules.
  - Case-based (e.g. KNN)
- **Pros**:
  - High interpretability
- **Cons:**
  - Low predictive power
  - Require a lot of human intervention

Predictions aren't enough... tell my why!

Logistic Regression Naive Bayes	Decision Trees Scorecards Inductive Rules	Random Forest GBDT AdaBoost	Deep Neural Networks

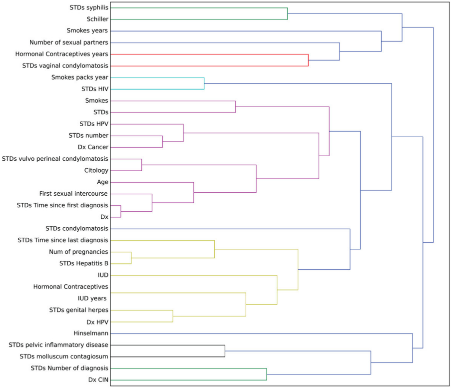
$\longleftrightarrow$

Low predictive power
High interpretability

High predictive power
Low interpretability

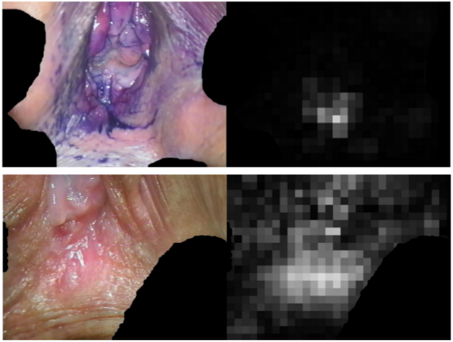
Interpretability and Deep Neural Networks

- **Sensitivity analysis (global - model):**



Fernandes, Kelwin, Davide Chicco, Jaime S. Cardoso, and Jessica Fernandes. "Supervised deep learning embeddings for the prediction of cervical cancer diagnosis." PeerJ Computer Science 4 (2018): e154.

- **Sensitivity analysis (local - instances):**



Fernandes, Kelwin, Jaime S. Cardoso, and Birgitte Schmidt Astrup. "A deep learning approach for the forensic evaluation of sexual assault." Pattern Analysis and Applications (2018): 1-12.

Interpretability and Deep Neural Networks

- Adversarial examples.
  - Adjust the observation to maximize the probability of observing the opponent class (backpropagation).
  - Gradients.
- Similarity in latent spaces (embeddings).
- Sparsity:
  - Complex decisions that depend on a small subset of variables

The holy grail of ML

- Fully automated
- High predictive power
- Limited data
- Fully interpretable

Machine Learning in Medical Applications: from fundamental topics to practical aspects

Kelwin Fernandes

INESC TEC

kelwinfc@gmail.com

June 12, 2018