Memorial Sloan Kettering Cancer Center

Machine Learning in Medical Physics

Lily Tang, Lei Xing, Hui Lin, and Timor Kadir

07/31/2017 1

Presentations

1. An overview of Machine Learning in Medical Physics – *Lily Tang*, Memorial Sloan Kettering Cancer Center.
2. Machine Learning for image analysis and treatment planning – *Lei Xing*, Stanford University.
3. Deep Learning in Medical Physics—lessons we learned – *Hui Lin*, Rensselaer Polytechnic Institute.
4. Deep-learned and practical: implementing machine learning techniques in the real world – *Timor Kadir*, Optellum Ltd, and Mirada Medical.

07/31/2017 2 Memorial Sloan Kettering Cancer Center

Memorial Sloan Kettering Cancer Center

An Overview of Machine Learning in Medical Physics

Lily Tang
Memorial Sloan Kettering Cancer Center
tangx@mskcc.org

07/31/2017 3

Outline

- A brief history of Machine Learning (ML)
- What is ML?
  - Types of the learning algorithms
  - No free lunch rule
  - How to evaluate and validate ML algorithms
- The future ML—Deep Learning
- Conclusion

A Brief History of ML

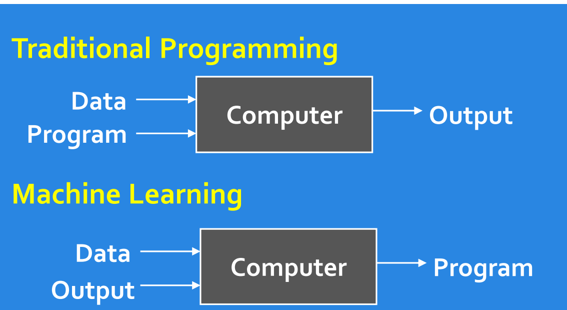
Year	Event
1763	The underpinnings of Bayes' Theorem.
1805	Least Squares described by Adrien-Marie Legendre.
1913	Markov Chains described by Andrey Markov.
1950	"Learning Machine" proposed by Alan Turing that could learn and become artificially intelligent.
1951	First Neural Network Machine built by Marvin Minsky and Dean Edmonds, able to learn.
1952	Machines playing Checkers. Arthur Samuel joined IBM's Poughkeepsie laboratory and began working on some of the very first machine learning programs, first creating programs that played checkers.
1957	The Perceptron invented by Frank Rosenblatt at Cornell Aeronautical Laboratory, which simulated the thought processes of the human brain.

A Brief History of ML (cont.)

Year	Event
1979	Stanford Cart developed by students at Stanford that can navigate and avoid obstacles in a room.
1985	NetTalk developed by Terry Sejnowski that learned to pronounce words the same way a baby does.
1997	IBM's Deep Blue beat the world champion at chess.
2006	Geoffrey Hinton coined the term "Deep Learning" to explain new algorithms that let computers "see" and distinguish objects and text in images and videos.
2014	Facebook published their work on DeepFace, a system that uses neural networks to identify faces with 97.35% accuracy—27% better than previous systems.
2015	Over 3,000 AI and Robotics researchers, endorsed by Stephen Hawking, Elon Musk and Steve Wozniak (among many others), signed an open letter warning of the danger of autonomous weapons which select and engage targets without human intervention.
2016	Google's AlphaGo beat a professional player at the Chinese board game Go—the world's most complex board game and is many times harder than chess.

What is Machine Learning (ML)?

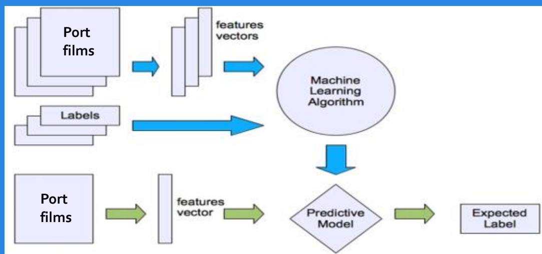
Machine Learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Machine Learning focuses on the development of computer programs that can access data and use it learn for themselves.



Types of Learning

1. Supervised learning
   - Training data includes desired outputs
   - Prediction, classification, etc.
2. Unsupervised learning
   - Training data does not include desired outputs
   - Clustering, probability distribution estimation, etc.
3. Reinforcement learning
   - Decision making (robot, chess machines, etc.)

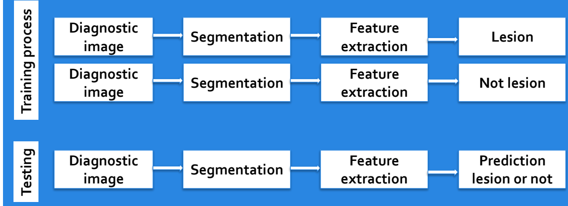
Supervised Learning



Picture credit: Chang & Ding

Supervised Learning Example 1

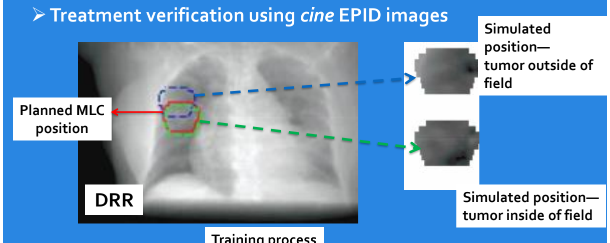
➢ Computerized detection of lesions in diagnostic images



Ref [1]

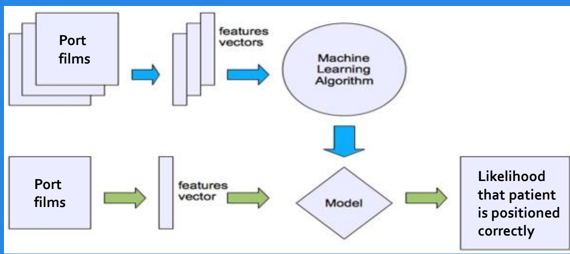
Supervised Learning Example 2

➢ Treatment verification using *cine* EPID images



Ref [2]

Unsupervised Learning

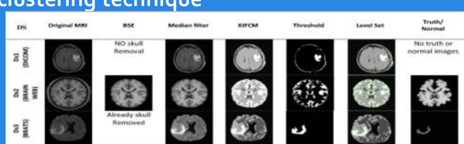


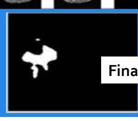
07/31/2017 13 Memorial Sloan Kettering Cancer Center

Picture credit: Chang & Ding

Unsupervised Learning Example

- Brain tumor segmentation based on a hybrid clustering technique



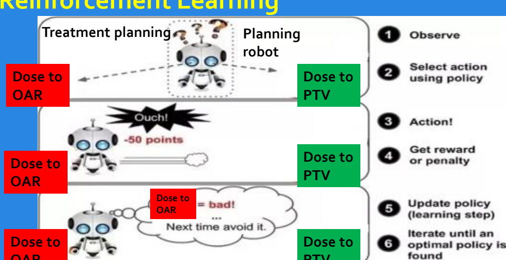


Final segmentation

Ref [3]

07/31/2017 Memorial Sloan Kettering Cancer Center

Reinforcement Learning



07/31/2017 15 Memorial Sloan Kettering Cancer Center

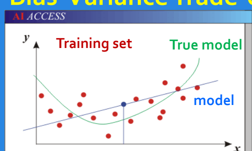
Picture credit: udacity.com

No Free Lunch Rule

- No model is inherently better than others: you need to make assumption to generalize

- Three kinds of error
  - Inherent: unavoidable
  - Bias: due to over-simplifications
  - Variance: due to inability to perfectly estimate parameters from limited data

Bias-Variance Trade-off

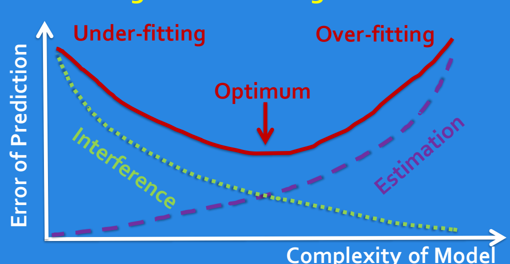


- Models with too few parameters are inaccurate because of a large bias (not enough flexibility).

- Models with too many parameters are inaccurate because of a large variance (too much sensitivity to the sample).

Slide credit: D. Hoiem

Under-fitting vs. Over-fitting



Evaluation of the ML Algorithms

- Confusion matrix

	Predicted class		
Actual class		True	False
	True	True Positive (TP)	False Negative (FN)
	False	False Positive (FP)	True Negative (TN)

Other Accuracy Measures

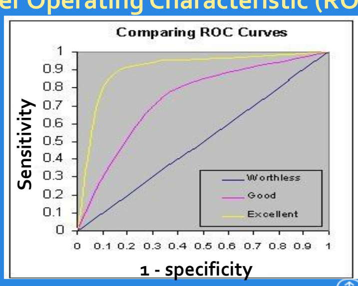
Model Accuracy = (TP + TN) / (TP + FP + TN + FN)

Misclassification Rate = (FP + FN) / (TP + FP + TN + FN)

Sensitivity = TP / (TP + FN)

Specificity = TN / (TN + FP)

Receiver Operating Characteristic (ROC)



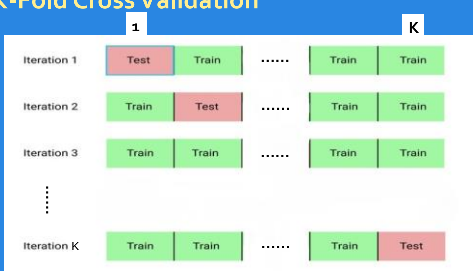
How to Validate the Model?

- Use a hold-out sample (in-sample testing)
- Validate using an sample not from the training period (out-of-time)
- Validate using an sample that is selected from a different population than that used to build the model (out-of-sample)

Evaluating Models

- How much data is available?
- Would be perfect if there was unlimited amount of data.
- In reality, we only have limited data—how to compromise?
  - Use hold-out testing sets
  - Cross validation
    - K-fold across validation
    - Leave-one-out validation

K-Fold Cross Validation



Validation Example

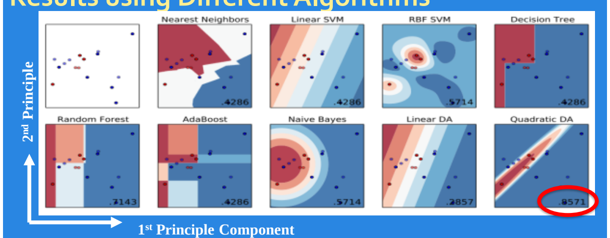
- Model based classification for optimal position selection for left-sided breast radiotherapy: DIBH, or prone
- 16 patients with free breathing, DIBH, and prone scans

Acquire supine free breathing scan → Extract features from FB CT scans → Labeled as DIBH/P based on dose criteria → PCA and train the statistical models → DIBH / Prone

- K = 5 and each experiment repeated 10 iterations
- Area Under Receiver Operating Characteristic (AUROC)

Ref [4]

Results using Different Algorithms

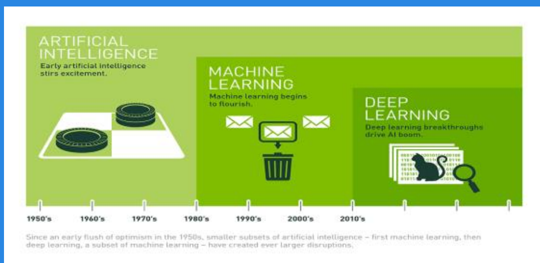


The red dots demonstrate DIBH positioning, and blue dots demonstrate prone positioning. The blue and red regions show the decision boundaries of each model. Numbers on the lower right corners are the AUROC.

Ongoing Validation

- Once your model have been deployed, you need to check if the model is still behaving as it was intended
  - Model accuracy
    - Run accuracy measurement on an ongoing basis
    - If statistics are deteriorating—need to re-balance the model or rebuild
  - Population stability
    - Compare the current distributions of the model inputs to the most recent model run
    - Report stability index

The Buzz Words—AI, ML, and DL



Picture credit: Nvidia

Conclusion

- We have visited the history of ML
- Introduced basic ML and examples in medical physics
- No free lunch rule—trade off is needed
- Learnt how to evaluate and validate ML algorithms

References

[1] I. El Naqa, R. Li, and M. Murphy, Machine Learning in Radiation Oncology Theory and Applications, *Springer*, 2015.
[2] X. Tang, T. Lin, and S. Jiang, A Feasibility Study of Treatment Verification Using EPID *cine* Images for Hypofractionated Lung Radiotherapy, *Physics in Medicine & Biology* 2009, 54, S1-S8.
[3] E. Abdel-Maksoud, M. Elmogy, and R. Al-Awadi, Brain Tumor Segmentation Based on a Hybrid Clustering Technique, *Egyptian Informatics Journal*, Vol. 16(1), pp.71-81, 2015.
[4] H. Lin, T. Liu, C. Shi, S. Petillion, I. Kindts, X. Tang, X. Xu, Model Based Classification for Optimal Position Selection for Left-Sided Breast Radiotherapy: Free Breathing, DIBH, Or Prone, *Medical Physics*, Vol. 43(6), pp.3529, 2016.


Memorial Sloan Kettering
Cancer Center
Thank you!
07/31/2017
31