



Some Advice on Applying Machine Learning in Practice

CS 760@UW-Madison





It's generalization that counts

- the fundamental goal of machine learning is to generalize beyond the instances in the training set
- you should rigorously measure generalization
- use a completely held-aside test set
- or use cross validation



It's generalization that counts

- but be careful not to let any information from test sets leak into training



- be careful about overfitting a data set, even when using cross validation



It's generalization that counts

- compare multiple learning approaches
- there is no single best approach





Data alone is not enough

- learning algorithms require inductive biases
 - smoothness
 - similar instances having similar classes
 - limited dependencies
 - limited complexity

Media Criticized For Biased Hometown Sports Reporting



VOLUME 241
NOVEMBER 20
50 CENTS
GLOBE WEST
NEWS, B2

The Boston Globe
MONDAY, FEBRUARY 4, 2002

CHAMPIONS!

Patriots win Super Bowl, 20-17, on Vinatieri's last-second kick

By Dan Dougherty
PHOTO BY AP/WIDEWORLD

NEW ORLEANS — The time, the half-kick, the
between opponent's legs. It added between the
winning end zone.
This time, there was no ill-fated penalty for too
many players on the field. No block. No bad
calls. No Charlie Brown kick.

Adam Vinatieri's last-second around-the-goal
kicked over the crossbar last night and gave the
New England Patriots a 20-17 victory over the St.
Louis Rams in what may have been the greatest
play of them all. Under the September's eye-
catcher sign, 1,200 miles from home, the Patriots
checked the scores and delivered Center Boston its
first professional sports championship since 1958.

On his way to becoming a Hall of Famer, Vinatieri
of the new era. All Tom Brady produced one final
touch to outplay the single play of 2001-02. The
winning opportunity consisted the number one
was by kicking the MVP trophy and becoming the
youngest winning QB in Super Bowl history.

"The way we did was that we were the better
team," said Brady. "Obviously, we were. It's what
happens when you believe in each other and don't
so many reasons why we're here."

Brady was barely 24 when he led the 27-140 yards,
one touchdown, no interceptions, but after the Rams
had the game, 17-17, with 1:30 remaining, he drove
the Patriots 12 yards in eight plays (including two
of interceptions), putting Vinatieri in position to win the
game.

...after four minutes of the overtime kickoff.

Fans revel in first title

By Michael Korda
PHOTO BY AP/WIDEWORLD
and David S. Greenberger
PHOTO BY AP/WIDEWORLD

T he knowledge you heard last night
was New England sports fans
celebrating the end of a 43-year
championship drought.




Data alone is not enough

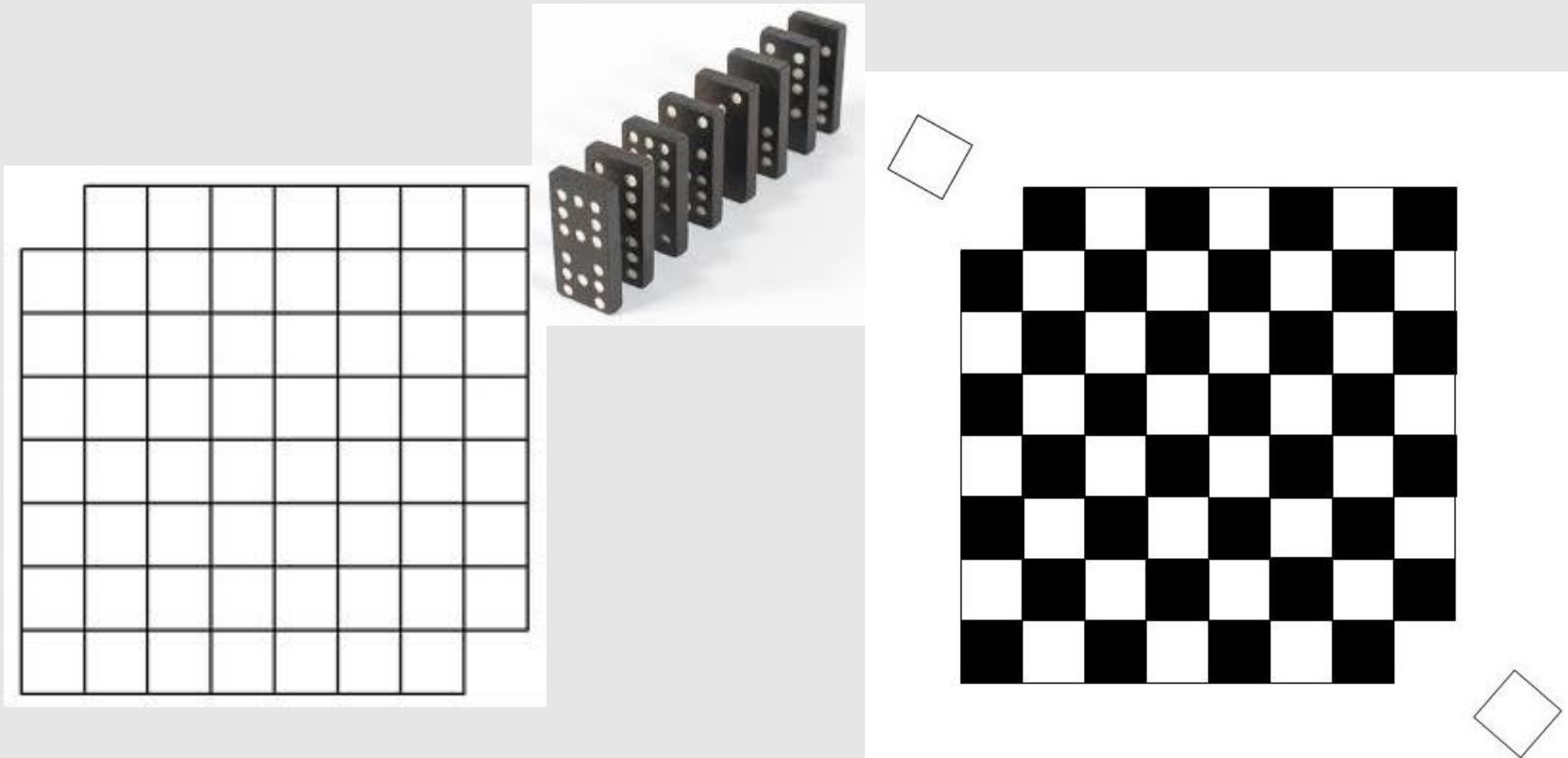


- when choosing a representation, consider what kinds of background knowledge are easily expressed in it
 - what makes instances similar → kernels
 - dependencies → graphical models
 - logical rules → inductive logic programming
 - etc.

The importance of representation



- each domino covers two squares
- can you cover the board with dominoes?



- the solution is more apparent when we change the representation



Feature engineering is key

- typically the most important factor in a learning task is the feature representation
- many independent features that correlate with class → learning is easy
- class is a complex function of features → learning is hard
- try to craft features that make apparent what might be most important for the task

Learn many models, not just one



- winning team and runner-up were both formed by merging multiple teams
- winning systems were ensembles with > 100 models
- combination of the the two winning systems was even more accurate

Learn many models, not just one



- the lesson is more general than the Netflix prize
- ensembles very often improve the accuracy of individual models

We may care more about the model than actually making predictions

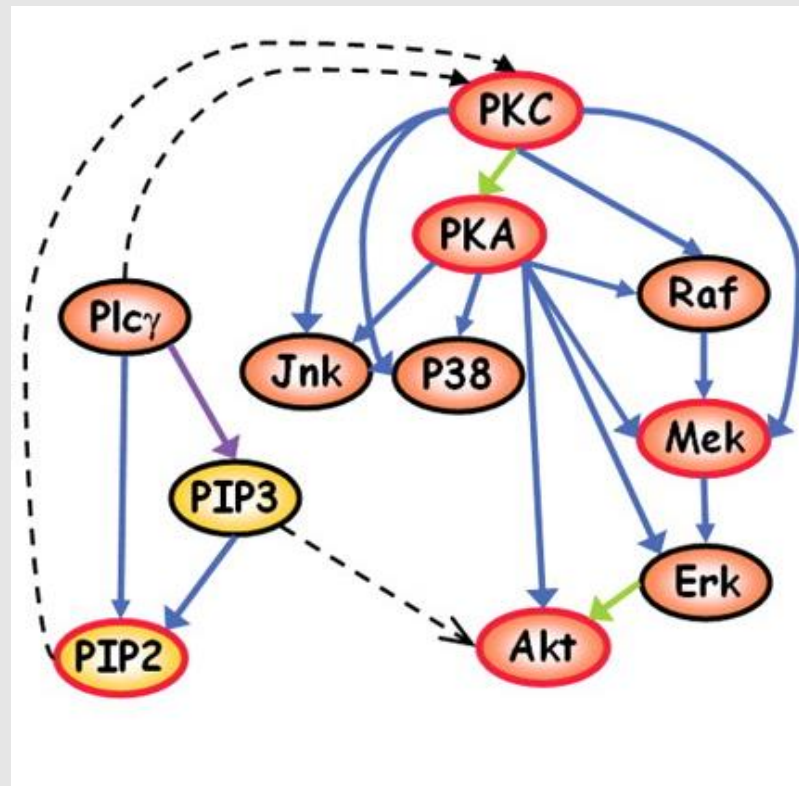


- two principal reasons for using machine learning
 1. to make predictions about test instances
 2. to gain insight into the problem domain
- for the former, a complicated black box may be okay
- for the latter, we want our models to be comprehensible to some degree

We may care more about the model than actually making predictions



- example: inferring Bayesian networks to represent intracellular networks [Sachs et al., *Science* 2005]



In many cases, we care about both



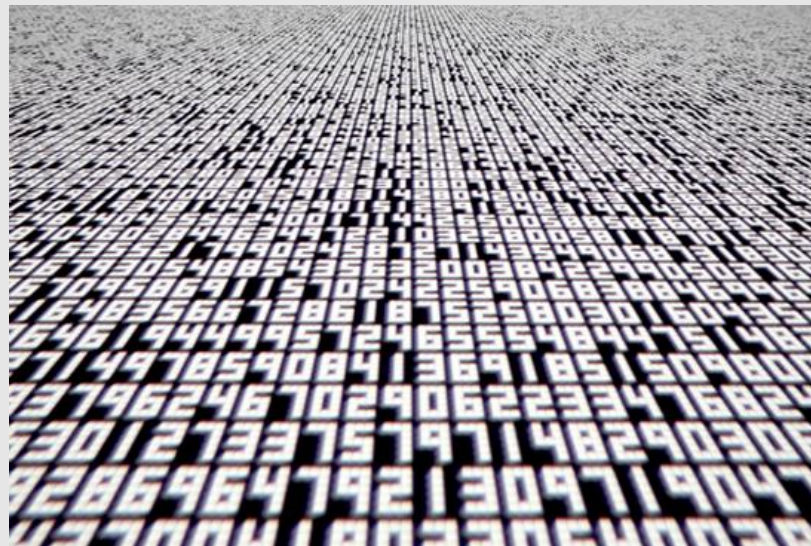
- example: predicting post-hospitalization VTE risk given patient histories [Kawaler et al., *AMIA* 2012]
 - want to identify patients at risk with high accuracy
 - want to identify previously unrecognized risk factors

Category	Risk Factor
Low Blood Volume	Furosemide
	Hypovolemia
	Hypo-osmolarity
	Posthemorrhagic Anemia
	Acute Renal Failure
Infection	E.Coli Infection
	Levofloxacin
	Cephalexin
Inflammation	High Alpha-1 Globulin Count
	Angina Pectoris
Immobilization	Pathologic Fracture of Vertebrae
Malnutrition	Protein Caloric Malnutrition

Theoretical guarantees are not what they seem



- PAC bounds are extremely loose
- asymptotic results tell us what happens when given infinite amounts of data – we don't usually have this
- learning theory results are generally
 - useful for understanding learning, driving algorithm design
 - not a criterion for practical decisions



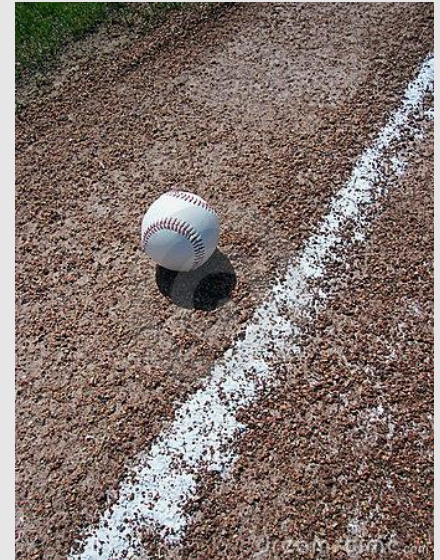
Do assumptions of algorithm hold?



- be sure to check the assumptions made by an approach/methodology against your problem domain
 - Are the instances *i.i.d.* or should we take into account dependencies among them?
 - When we divide a data set into training/test sets, is the division representative of how the learner will be used in practice?
 - etc.
- questioning the assumptions of standard approaches sometimes results in new paradigms
 - active learning
 - multiple-instance learning
 - etc.

Compare against reasonable baselines

- Empirically determine whether fancy ML methods have value by comparing against
 - simple predictors (e.g. tomorrow's weather will be the same as today's)
 - standard predictors in use
 - individual features





THANK YOU

Some of the slides in these lectures have been adapted/borrowed from materials developed by Mark Craven, David Page, Jude Shavlik, Tom Mitchell, Nina Balcan, Elad Hazan, Tom Dietterich, and Pedro Domingos.

