Incompatibilities(?) between PACBayes and Exploration

John Langford (Yahoo! Research)

PAC Bayes Workshop

March 22, 2010

What is a PAC-Bayes bound?

What is a PAC-Bayes bound?

1. Tightness: Tight sample complexity bounds.

What is a PAC-Bayes bound?

1. Tightness: Tight sample complexity bounds.
2. Luckiness: Variable competition based on a prior.

What is a PAC-Bayes bound?

1. Tightness: Tight sample complexity bounds.
2. Luckiness: Variable competition based on a prior.
3. Indifference: You don't pay for irrelevant decisions.

Basic claim: achieving (2) and (3) are inherently problematic in situations with exploration.

Outline

1. Supervised Learning and PAC-Bayes Review
2. Active Learning and PAC-Bayes
3. Contextual Bandits and PAC-Bayes

Supervised Learning Setting

Repeatedly:

1. The world reveals features x.
2. A learning algorithm chooses a label $\hat{y} \in \{0, 1\}$.
3. The world reveals a label $y \in \{0, 1\}$.

Goal: Compete with hypothesis class $H = \{h : X \to Y\}$.

Typical Algorithm and Theorem

Let $e(h, D) = \Pr_{(x,y)\sim D}(h(x) \neq y)$ and
$e(h, S) = \frac{1}{|S|} \sum_{(x,y)\in S} I(h(x) \neq y)$

Typical Algorithm and Theorem

Let $e(h, D) = \Pr_{(x,y)\sim D}(h(x) \neq y)$ and $e(h, S) = \frac{1}{|S|} \sum_{(x,y)\in S} I(h(x) \neq y)$

Algorithm: ERM

1. Observe x
2. Let $h_{\text{min}} = \arg\min_{h\in H} e(h, S)$
3. return $\hat{y} = h_{\text{min}}(x)$

Typical Algorithm and Theorem

Let $e(h, D) = \Pr_{(x,y)\sim D}(h(x) \neq y)$ and
$e(h, S) = \frac{1}{|S|} \sum_{(x,y)\in S} I(h(x) \neq y)$

Algorithm: ERM

1. Observe x
2. Let $h_{\text{min}} = \arg\min_{h\in H} e(h, S)$
3. return $\hat{y} = h_{\text{min}}(x)$

Theorem: For all IID distributions D, for all hypothesis sets H, over T timesteps with probability $1 - \delta$

$$e(\text{argmin}_S, D) - e(h^*, D) \leq O\left(\sqrt{\frac{\ln|H| + \ln\frac{1}{\delta}}{T}}\right)$$

where $\text{regret} = e(\text{argmin}_S, D) - e(h^*, D)$

Luckiness and Indifference for Supervised Learning

Luckiness: (Occam's Razor) For all $D, H, P(h)$, with probability $1-\delta$:

$$e(\mathrm{argmin}_S, D) - e(h^*, D) \leq O\left(\sqrt{\frac{\ln \frac{1}{P(h^*)} + \ln \frac{1}{\delta}}{T}}\right)$$

$\Rightarrow \mathrm{regret}(H_1 \cup H_2) \leq O(1) + \max\{\mathrm{regret}(H_1), \mathrm{regret}(H_2)\}$

Luckiness and Indifference for Supervised Learning

Luckiness: (Occam's Razor) For all $D, H, P(h)$, with probability $1-\delta$:

$$e(\text{argmin}_S, D) - e(h^*, D) \leq O\left(\sqrt{\frac{\ln \frac{1}{P(h^*)} + \ln \frac{1}{\delta}}{T}}\right)$$

$$\Rightarrow \text{regret}(H_1 \cup H_2) \leq O(1) + \max\{\text{regret}(H_1), \text{regret}(H_2)\}$$

Indifference: (PAC-Bayes) For all D, H, with probability $1-\delta$ for all $Q(h)$:

$$e(Q, S) - e(Q, D) \leq O\left(\sqrt{\frac{\ln |H| - H(Q) + \ln \frac{1}{\delta}}{T}}\right)$$

Let H_2 consist of h satisfying $e(h, D) = \min_{h' \in H_1} e(h', D)$.
$\Rightarrow \text{regret}(H_1 \cup H_2) \leq \text{regret}(H_1)$ (or less!)

Outline

1. Supervised Learning and PAC-Bayes Review
2. Active Learning and PAC-Bayes
3. Contextual Bandits and PAC-Bayes

Active Learning Setting

Repeatedly:

1. The world reveals features x.
2. A learning algorithm chooses an action $\hat{y} \in \{0, 1\}$.
3. The world reveals a label $y \in \{0, 1\}$ if requested by the algorithm.

Goal: Compete with hypothesis class $H = \{h : X \to Y\}$ while minimizing label complexity.

Typical Algorithm and Analysis

Keep track of a version space H_S which is initially H.

1. Observe an x,
2. Predict according to $\arg\min_{h \in H_S} e(h, S)$.
3. if $\exists h, h' \in H_S$ satisfying $h(x) \neq h'(x)$.
   1. ask for the label
   2. Use a sample complexity bound to remove all h provably not optimal from H_S.

Typical Algorithm and Analysis

Keep track of a version space H_S which is initially H.

1. Observe an x,
2. Predict according to $\arg\min_{h \in H_S} e(h, S)$.
3. if $\exists h, h' \in H_S$ satisfying $h(x) \neq h'(x)$.
    1. ask for the label
    2. Use a sample complexity bound to remove all h provably not optimal from H_S.

Theorem: For all IID distributions D, for all hypothesis sets H, over T timesteps with probability $1 - \delta$:

$$\text{Active learning regret} = \text{O(Supervised Regret)}$$

and

$$\#labels \leq O\left(\theta \ln |H| (\ln T)(\ln T + \ln 1/\delta)\right)$$

for $e(h^*, D)$ small and $\theta =$ "disagreement coefficient"

The disagreement coefficient

Let $H_\epsilon = \{h \in H : e(h, D) \leq e(h^*, D) + \epsilon\}$
Let $X_\epsilon = \{x \in X : h(x) \neq h^*(x)\}$
Disagreement coefficient $= \max_\epsilon \frac{D(X_\epsilon)}{\epsilon}$

The disagreement coefficient

Let $H_\epsilon = \{h \in H : e(h, D) \leq e(h^*, D) + \epsilon\}$
Let $X_\epsilon = \{x \in X : h(x) \neq h^*(x)\}$
Disagreement coefficient $= \max_\epsilon \frac{D(X_\epsilon)}{\epsilon}$

Two hypotheses h_1, h_2, any data distribution.

The disagreement coefficient

Let $H_\epsilon = \{h \in H : e(h, D) \leq e(h^*, D) + \epsilon\}$
Let $X_\epsilon = \{x \in X : h(x) \neq h^*(x)\}$
Disagreement coefficient $= \max_\epsilon \frac{D(X_\epsilon)}{\epsilon}$

Two hypotheses h_1, h_2, any data distribution.

$$\theta = 1.$$

Thresholds in R, any data distribution.

The disagreement coefficient

Let $H_\epsilon = \{h \in H : e(h, D) \leq e(h^*, D) + \epsilon\}$
Let $X_\epsilon = \{x \in X : h(x) \neq h^*(x)\}$
Disagreement coefficient $= \max_\epsilon \frac{D(X_\epsilon)}{\epsilon}$

Two hypotheses h_1, h_2, any data distribution.

$$\theta = 1.$$

Thresholds in R, any data distribution.

$$\theta = 2.$$

Linear separators through the origin in R^d, uniform data distribution.

The disagreement coefficient

Let $H_\epsilon = \{h \in H : e(h, D) \leq e(h^*, D) + \epsilon\}$
Let $X_\epsilon = \{x \in X : h(x) \neq h^*(x)\}$
Disagreement coefficient $= \max_\epsilon \frac{D(X_\epsilon)}{\epsilon}$

Two hypotheses h_1, h_2, any data distribution.

$$\theta = 1.$$

Thresholds in R, any data distribution.

$$\theta = 2.$$

Linear separators through the origin in R^d, uniform data distribution.

$$\theta \leq \sqrt{d}.$$

Luckiness: ??

Luckiness and Indifference for Active Learning

Luckiness: ??

$\theta(H_1 \cup H_2, D) \leq \theta(H_1, D) + \theta(H_2, D) + O(1)$
When learning on $H_1 \cup H_2$ label complexities *add* in the worst case.
$\Rightarrow$ dealing with a prior is very difficult.

Luckiness and Indifference for Active Learning

Luckiness: ??

$\theta(H_1 \cup H_2, D) \leq \theta(H_1, D) + \theta(H_2, D) + O(1)$
When learning on $H_1 \cup H_2$ label complexities *add* in the worst case.
$\Rightarrow$ dealing with a prior is very difficult.

Indifference: ??

Luckiness and Indifference for Active Learning

Luckiness: ??

$\theta(H_1 \cup H_2, D) \leq \theta(H_1, D) + \theta(H_2, D) + O(1)$
When learning on $H_1 \cup H_2$ label complexities *add* in the worst case.
$\Rightarrow$ dealing with a prior is very difficult.

Indifference: ??

Let H_2 consist of h satisfying $e(h, D) = \min_{h' \in H_1} e(h', D)$.
$\Rightarrow \theta(H_1 \cup H_2, D) \leq \theta(H_1, D) + \frac{e(h^*, D)}{\epsilon}|H_2|$
Extra good hypotheses can *hurt*.

So, PAC-Bayes appears incompatible with this style of active learning.

Some objections you might have

Some objections you might have

Cardinal sin! You compare upper bound to upper bound rather than upper bound to lower bound!

Some objections you might have

Cardinal sin! You compare upper bound to upper bound rather than upper bound to lower bound!

Sure, but there are some lower bounds involving disagreement.

Some objections you might have

Cardinal sin! You compare upper bound to upper bound rather than upper bound to lower bound!
Sure, but there are some lower bounds involving disagreement.
Should we care about active learning? The analysis looks rather finicky/loose/unclean.

Some objections you might have

Cardinal sin! You compare upper bound to upper bound rather than upper bound to lower bound!

Sure, but there are some lower bounds involving disagreement.

Should we care about active learning? The analysis looks rather finicky/loose/unclean.

1. Lots of people care, empirically.
2. The theory is starting to yield useful algorithms (see IWAL paper).
3. Maybe not, but that's why I brought another exploration setting.

Outline

1. Supervised Learning and PAC-Bayes Review
2. Active Learning and PAC-Bayes
3. Contextual Bandits and PAC-Bayes

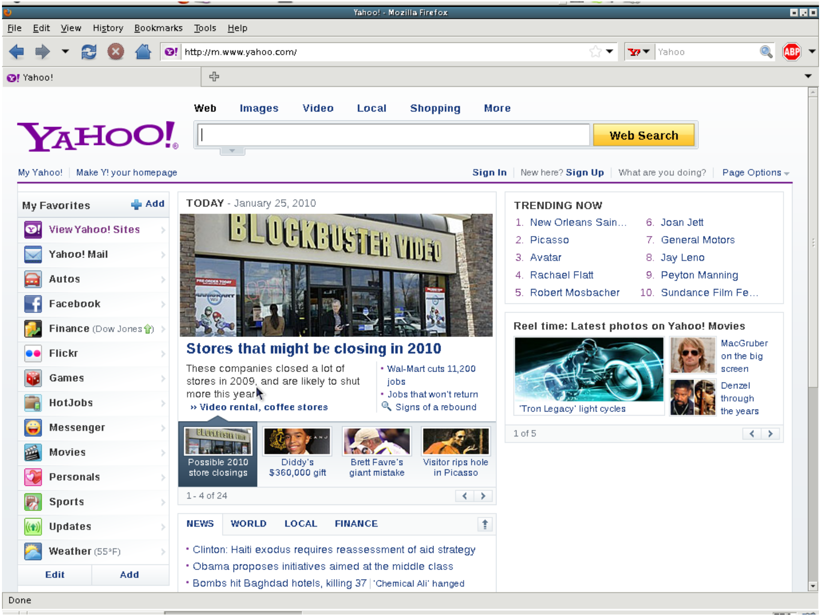
Contextual Bandits Setting

Repeatedly:

1. The world reveals features x.
2. A learning algorithm chooses an label $\hat{y} \in \{0, 1\}$.
3. The world reveals reward $r \in [0, 1]$ for action a.

Goal: Compete with hypothesis class $H = \{h : X \rightarrow Y\}$.iu

This setting is very easy to motivate at Y!


Yahoo! - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://m.www.yahoo.com/
Yahoo
Yahoo!
Web Images Video Local Shopping More
YAHOO!
Web Search
My Yahoo! | Make Y! your homepage
Sign In | New here? Sign Up | What are you doing? | Page Options
My Favorites
Add
View Yahoo! Sites
Yahoo! Mail
Autos
Facebook
Finance (Dow Jones)
Flickr
Games
HotJobs
Messenger
Movies
Personals
Sports
Updates
Weather (55°F)
Edit
Add
TODAY - January 25, 2010
BLOCKBUSTER VIDEO
Stores that might be closing in 2010
These companies closed a lot of stores in 2009, and are likely to shut more this year.
» Video rental, coffee stores
• Wal-Mart cuts 11,200 jobs
• Jobs that won't return
Signs of a rebound
Possible 2010 store closings
Diddy's $360,000 gift
Brett Favre's giant mistake
Visitor rips hole in Picasso
1 - 4 of 24
NEWS WORLD LOCAL FINANCE
• Clinton: Haiti exodus requires reassessment of aid strategy
• Obama proposes initiatives aimed at the middle class
• Bombs hit Baghdad hotels, killing 37 | 'Chemical Ali' hanged
TRENDING NOW
1. New Orleans Sain...
2. Picasso
3. Avatar
4. Rachael Flatt
5. Robert Mosbacher
6. Joan Jett
7. General Motors
8. Jay Leno
9. Peyton Manning
10. Sundance Film Fe...
Reel time: Latest photos on Yahoo! Movies
'Tron Legacy' light cycles
MacGruber on the big screen
Denzel through the years
1 of 5
Done

A Simple Algorithm and theorem

Keep track of instantaneous regret R and observations $S = (x, y, r_y)^*$.

Let $e'(h, S) = \sum_{(x,y,r_y,p_y)\in S} 2r_y I(h(x) = y)$

A Simple Algorithm and theorem

Keep track of instantaneous regret R and observations $S = (x, y, r_y)^*$.

Let $e'(h, S) = \sum_{(x,y,r_y,p_y)\in S} 2r_y I(h(x) = y)$

Algorith Epoch-Greedy:

1. With probability $1 - R$ predict according to $\arg\min_{h\in H_S} e'(h, S)$.
2. Otherwise choose an action at random.
3. Observe the reward r_y.
4. If the random action was taken, update S and R.

A Simple Algorithm and theorem

Keep track of instantaneous regret R and observations $S = (x, y, r_y)^*$.

Let $e'(h, S) = \sum_{(x,y,r_y,p_y)\in S} 2r_y I(h(x) = y)$

Algorith Epoch-Greedy:

1. With probability $1 - R$ predict according to $\arg\min_{h\in H_S} e'(h, S)$.
2. Otherwise choose an action at random.
3. Observe the reward r_y.
4. If the random action was taken, update S and R.

Theorem: For all IID distributions D, for all hypothesis sets H, over T timesteps with probability $1 - \delta$

$$\text{average regret} \leq O\left(\left(\frac{\ln|H| + \ln\frac{1}{\delta}}{T}\right)^{1/3}\right)$$

Note: EXP4P is a more complicated algorithm replacing $1/3$ with $1/2$.

Computation of R

The essential idea is to use a sample complexity bound.
Any will do.
The result comes from applying the sqrt-form Occam's Razor bound with a uniform prior.

Luckiness and Indifference for Contextual Bandit

Luckiness: ??

Luckiness and Indifference for Contextual Bandit

Luckiness: ??

Plugging in a nonuniform prior $\Rightarrow R \simeq \max_h \sqrt{\frac{\ln 1/P(h)}{T}}$ which is untenable. (Problem is shared by EXP4P.)
In the worst case, average regrets can add when combining two hypothesis spaces.

Luckiness and Indifference for Contextual Bandit

Luckiness: ??

Plugging in a nonuniform prior $\Rightarrow R \simeq \max_h \sqrt{\frac{\ln 1/P(h)}{T}}$ which is untenable. (Problem is shared by EXP4P.)
In the worst case, average regrets can add when combining two hypothesis spaces.

Indifference: ??

Luckiness and Indifference for Contextual Bandit

Luckiness: ??

Plugging in a nonuniform prior $\Rightarrow R \simeq \max_h \sqrt{\frac{\ln 1/P(h)}{T}}$ which is untenable. (Problem is shared by EXP4P.)
In the worst case, average regrets can add when combining two hypothesis spaces.

Indifference: ??

Indifference works out. Let H_2 consist of h satisfying $e'(h, D) = \min_{h' \in H_1} e(h', D)$. Then, the argmin can be replaced by randomization Q over a good set. (Even more clear with EXP4P.)

Conclusion

Something has to give in active learning for a PAC-Bayes benefit.
Perhaps we need to use unsupervised data as a resource?

Conclusion

Something has to give in active learning for a PAC-Bayes benefit. Perhaps we need to use unsupervised data as a resource?

For Contextual Bandits, indifference may work out ok, but there are substantial problems with luckiness.

Conclusion

Something has to give in active learning for a PAC-Bayes benefit. Perhaps we need to use unsupervised data as a resource?

For Contextual Bandits, indifference may work out ok, but there are substantial problems with luckiness.

More discussion about Active Learning and Exploration at http://hunch.net

Bibliography

[EXP4P] Alina Beygelzimer, John Langford, Lihong Li, Lev Reyzin, Robert E. Schapire, An Optimal High Probability Algorithm for the Contextual Bandit Problem 2010.

[IWAL] Alina Beygelzimer, Sanjoy Dasgupta, and John Langford, Importance Weighted Active Learning, ICML 2009.

[Active Learning] Sanjoy Dasgupta, Daniel Hsu, and Claire Monteleoni. A general agnostic active learning algorithm. NIPS 2007.

[Epoch-Greedy] John Langford and Tong Zhang The Epoch-Greedy Algorithm for Contextual Multi-armed Bandits NIPS 2007.