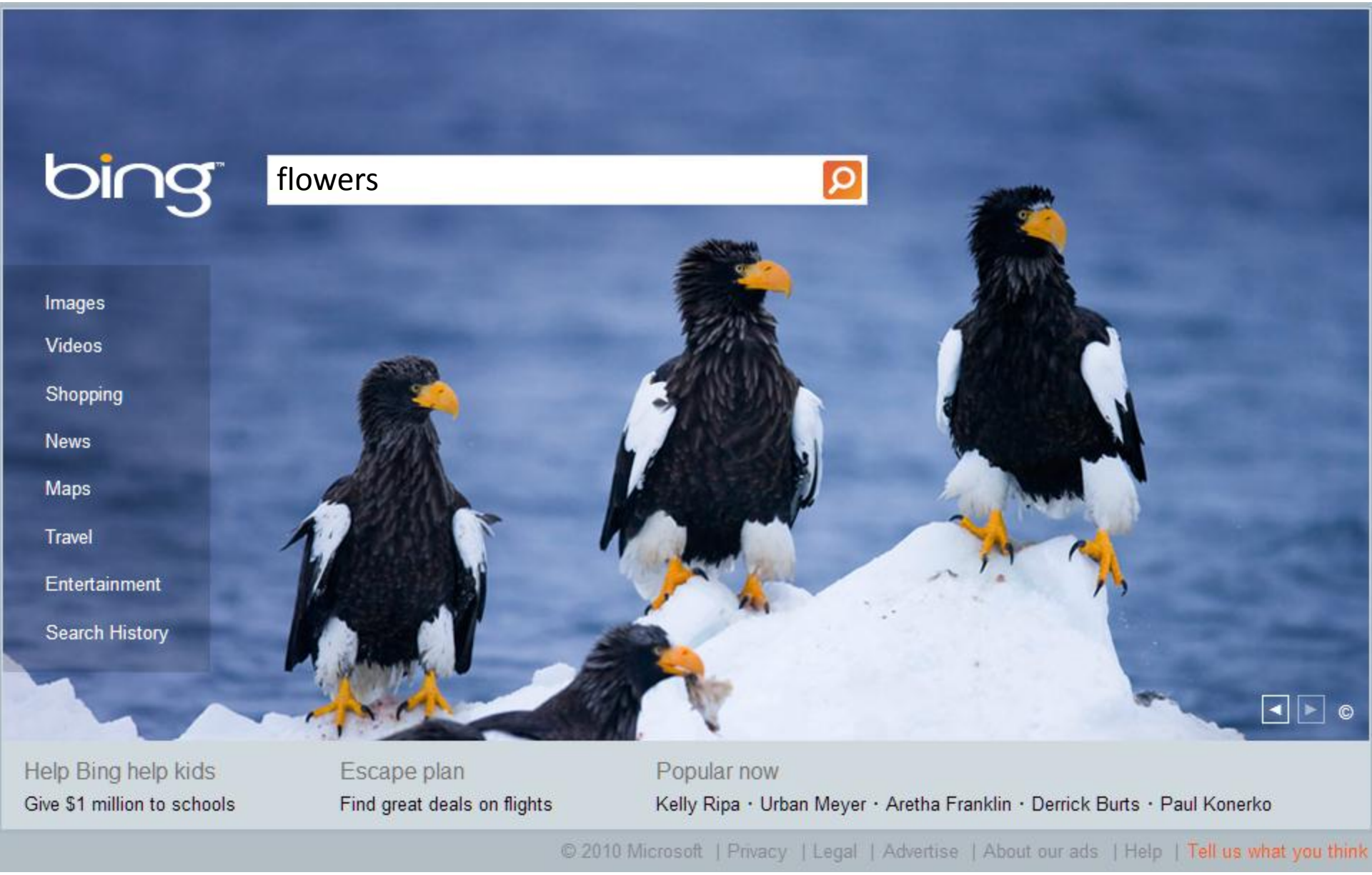
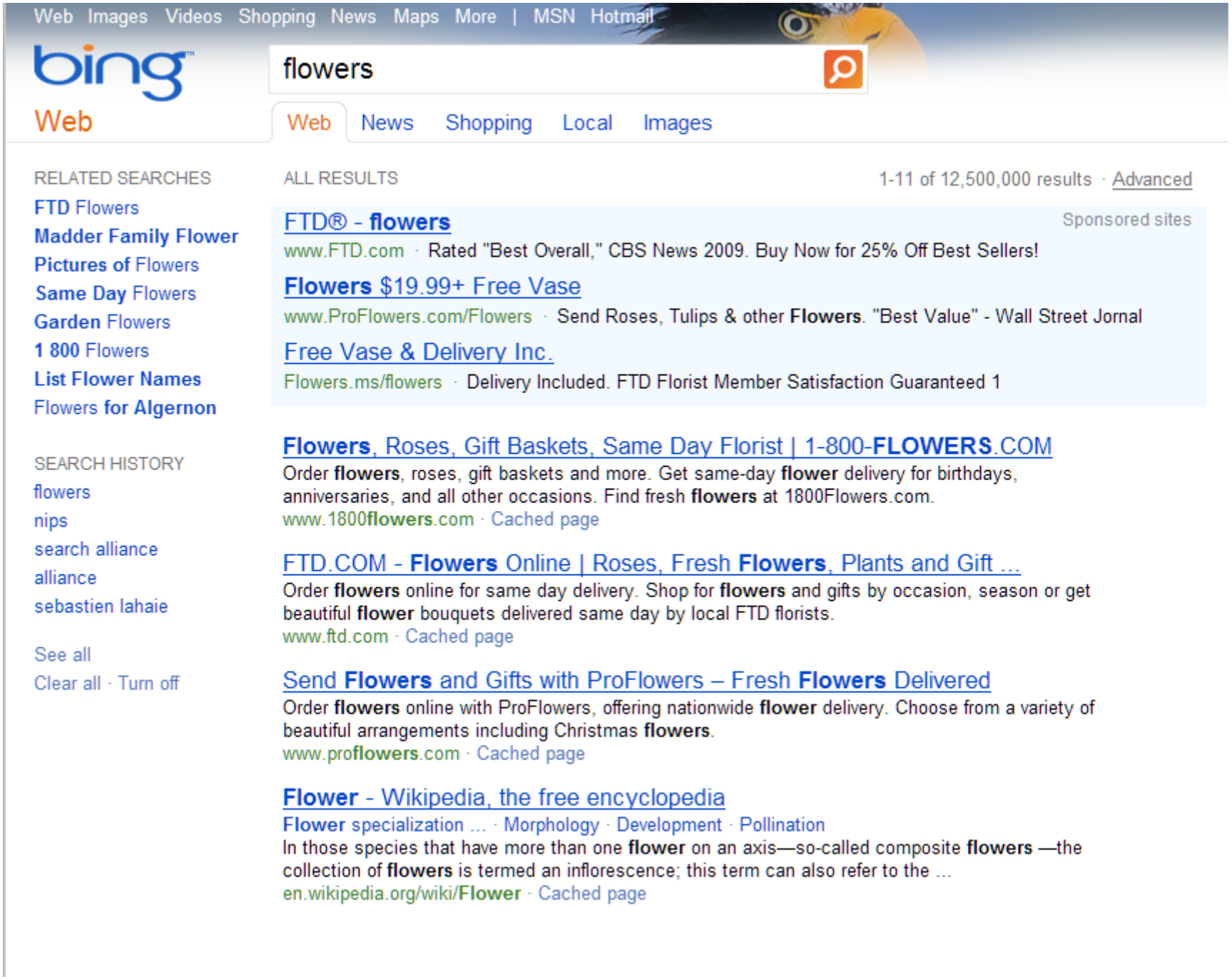
Click Prediction with adPredictor at Microsoft Advertising

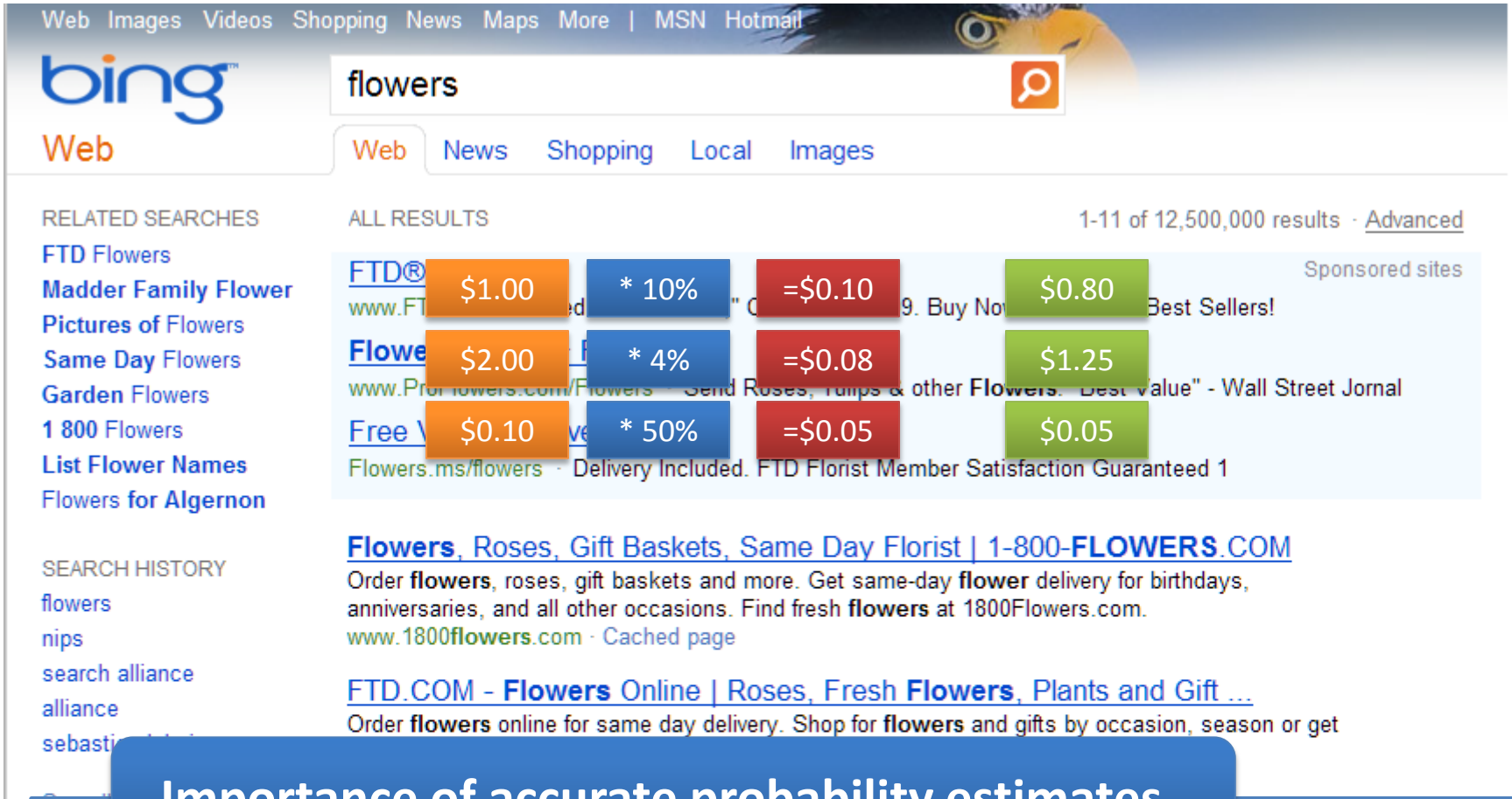
Joaquin Quiñonero Candela
Thore Graepel
Ralf Herbrich
Thomas Borchert
Microsoft Research & Microsoft adCenter

Microsoft + Yahoo! = 1/3 US search market

adPredictor predicts probability of click on ads for Microsoft Bing and Yahoo! search engines


bing
flowers
Images
Videos
Shopping
News
Maps
Travel
Entertainment
Search History
Help Bing help kids
Give $1 million to schools
Escape plan
Find great deals on flights
Popular now
Kelly Ripa · Urban Meyer · Aretha Franklin · Derrick Burts · Paul Konerko
© 2010 Microsoft | Privacy | Legal | Advertise | About our ads | Help | Tell us what you think


Web Images Videos Shopping News Maps More | MSN Hotmail
bing
flowers
Web
Web News Shopping Local Images
RELATED SEARCHES
FTD Flowers
Madder Family Flower
Pictures of Flowers
Same Day Flowers
Garden Flowers
1 800 Flowers
List Flower Names
Flowers for Algernon
SEARCH HISTORY
flowers
nips
search alliance
alliance
sebastien lahaie
See all
Clear all · Turn off
ALL RESULTS
1-11 of 12,500,000 results · Advanced
Sponsored sites
FTD® - flowers
www.FTD.com · Rated "Best Overall," CBS News 2009. Buy Now for 25% Off Best Sellers!
Flowers $19.99+ Free Vase
www.ProFlowers.com/Flowers · Send Roses, Tulips & other Flowers. "Best Value" - Wall Street Jornal
Free Vase & Delivery Inc.
Flowers.ms/flowers · Delivery Included. FTD Florist Member Satisfaction Guaranteed 1
Flowers, Roses, Gift Baskets, Same Day Florist | 1-800-FLOWERS.COM
Order flowers, roses, gift baskets and more. Get same-day flower delivery for birthdays, anniversaries, and all other occasions. Find fresh flowers at 1800Flowers.com.
www.1800flowers.com · Cached page
FTD.COM - Flowers Online | Roses, Fresh Flowers, Plants and Gift ...
Order flowers online for same day delivery. Shop for flowers and gifts by occasion, season or get beautiful flower bouquets delivered same day by local FTD florists.
www.ftd.com · Cached page
Send Flowers and Gifts with ProFlowers – Fresh Flowers Delivered
Order flowers online with ProFlowers, offering nationwide flower delivery. Choose from a variety of beautiful arrangements including Christmas flowers.
www.proflowers.com · Cached page
Flower - Wikipedia, the free encyclopedia
Flower specialization ... · Morphology · Development · Pollination
In those species that have more than one flower on an axis—so-called composite flowers —the collection of flowers is termed an inflorescence; this term can also refer to the ...
en.wikipedia.org/wiki/Flower · Cached page

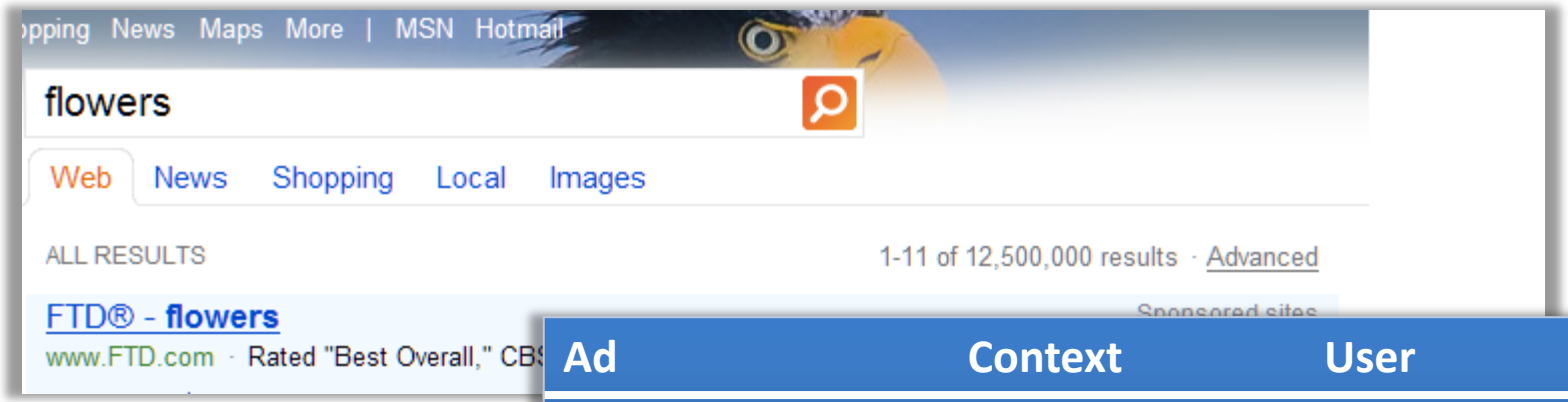


Importance of accurate probability estimates

- Efficient use of ad space
- Increased user satisfaction by better targeting
- Increased revenue by showing ads with high click-thru rate

Over-simplified ranking function: this is not what is used in practice

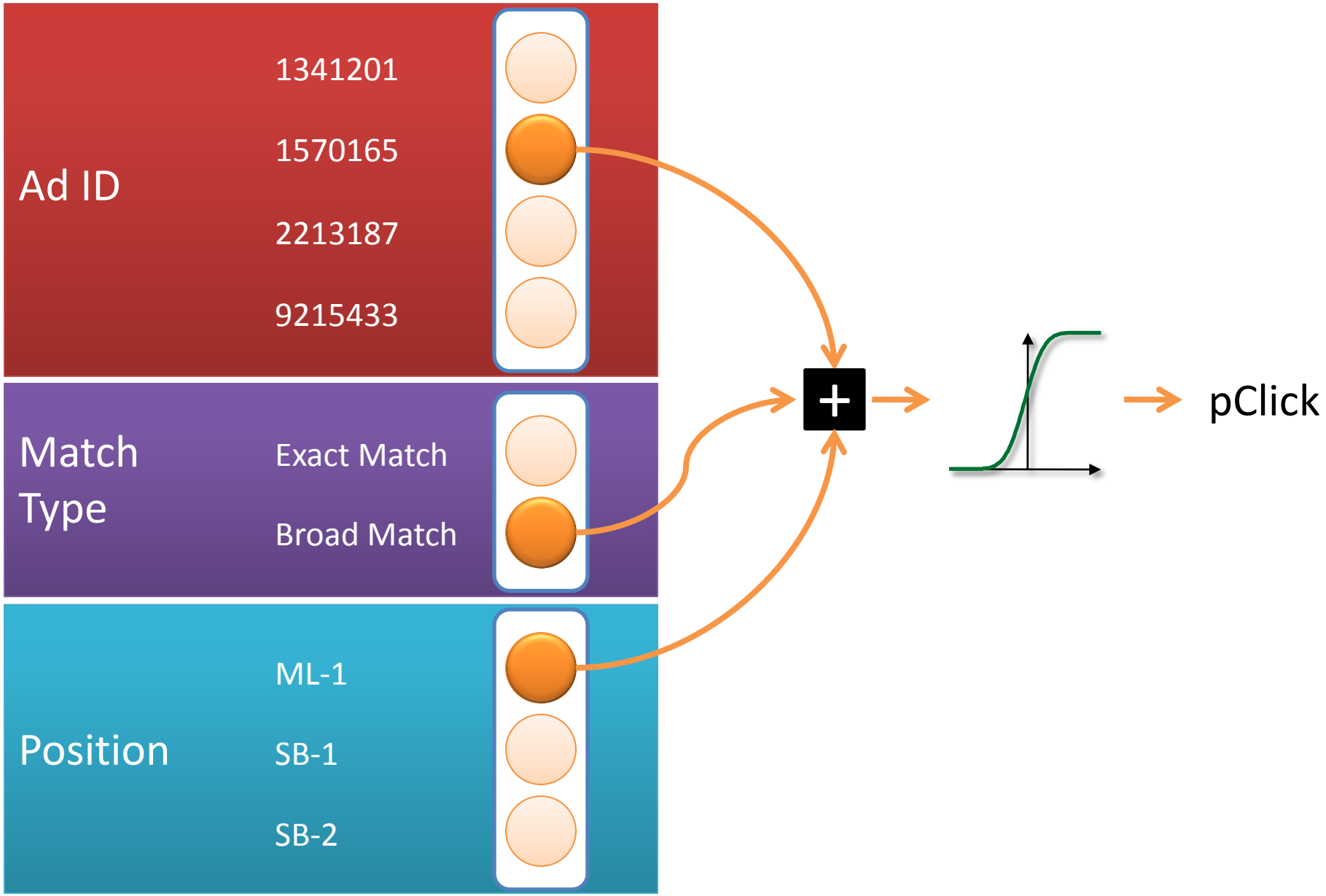
Impression Level Predictions



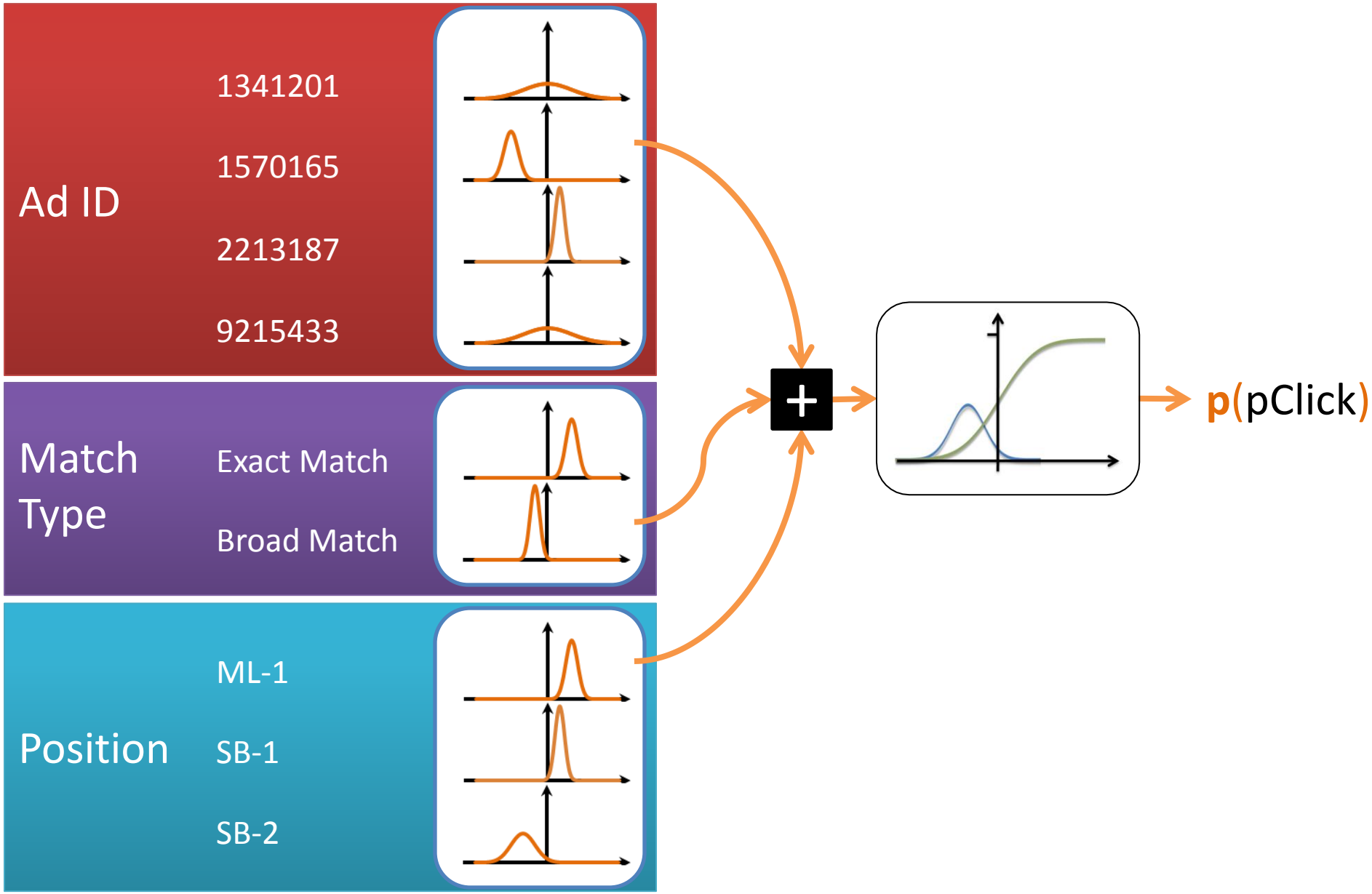
Ad	Context	User
AdId	Display Position	Query
Match Type	Date & Time	… and many more

- Sparse binary input features (many 10s of them)
- Some high cardinality (~100M), some low (<10)

Sparse Linear Probit Regression



Uncertainty: A Bayesian Treatment



A Linear Probit Model

- Notation

  $y = 1$ if click — $\boldsymbol{w}$ is the vector of all weights

  $y = -1$ if non-click — $\boldsymbol{x}$ is a sparse binary input vector

- Generalised linear model with weights vector $\boldsymbol{w}$:

$$p(y|\boldsymbol{x}, \boldsymbol{w}) := \Phi\left(\frac{y \cdot \boldsymbol{w}^T \boldsymbol{x}}{\beta}\right)$$

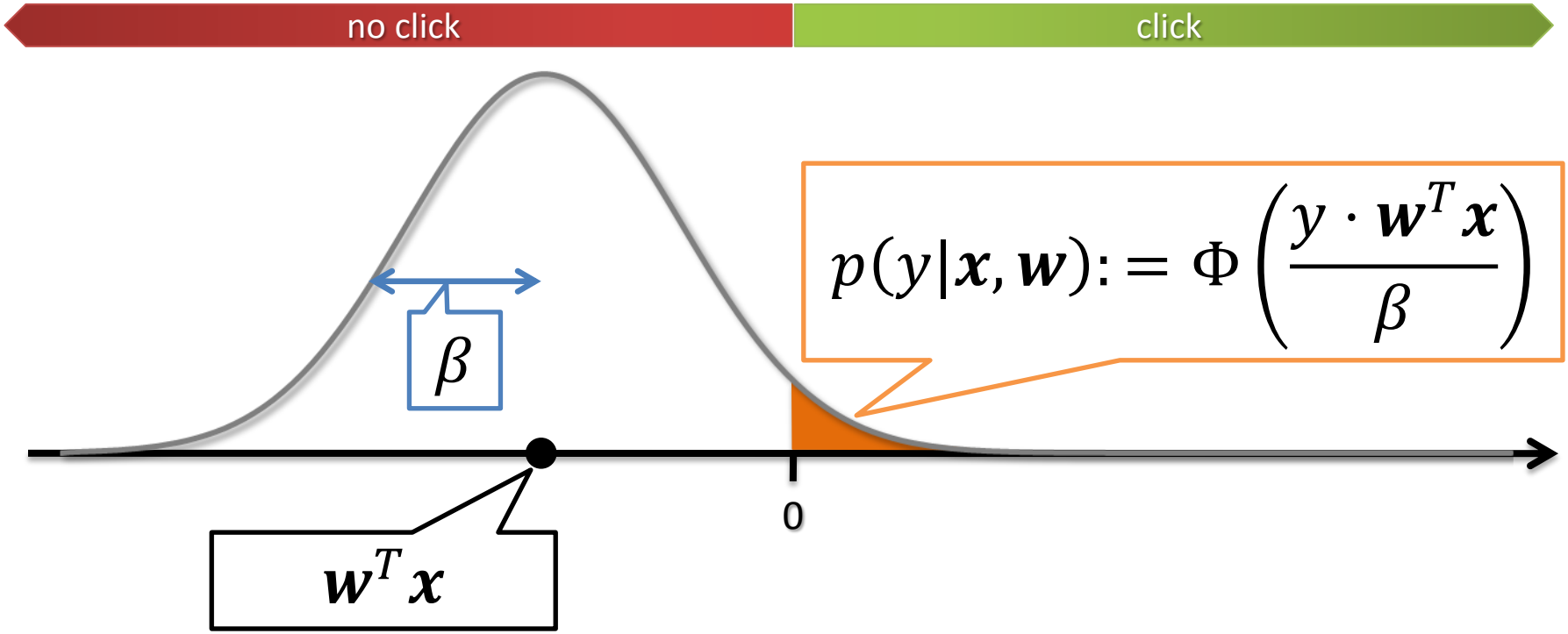
- Inverse link function is the probit function:

$$\Phi(t) := \int_{-\infty}^{t} \mathcal{N}(z; 0,1)\, dz$$

β controls the steepness: it corresponds to the standard deviation of additive zero mean noise.

Observation Noise

(Assume Known Noiseless Weights)



Think of $\boldsymbol{x}$ as indicator variables that select weights: we will soon remove $\boldsymbol{x}$ from the notation

Example = $x = [1; 0; 0; 0; 1; 0; \dots; 0; 1]$

Uncertainty About the Weights

A Bayesian Treatment

- Factorizing Gaussian prior over the weights:

$$p(\boldsymbol{w}) = \prod_{i=1}^{N} \mathcal{N}\left(w_i; \mu_i, \sigma_i^2\right)$$

- Given $p(y|\boldsymbol{x}, \boldsymbol{w})$ the **posterior** is given by:

$$p(\boldsymbol{w}|\boldsymbol{x}, y) = \frac{p(y|\boldsymbol{x}, \boldsymbol{w}) \cdot p(\boldsymbol{w})}{\int p(y|\boldsymbol{x}, \boldsymbol{w}) \cdot p(\boldsymbol{w}) \cdot \mathrm{d}\boldsymbol{w}}$$

Problem: This posterior cannot be represented compactly nor calculated in closed form

Desiderata and Approximations

We want

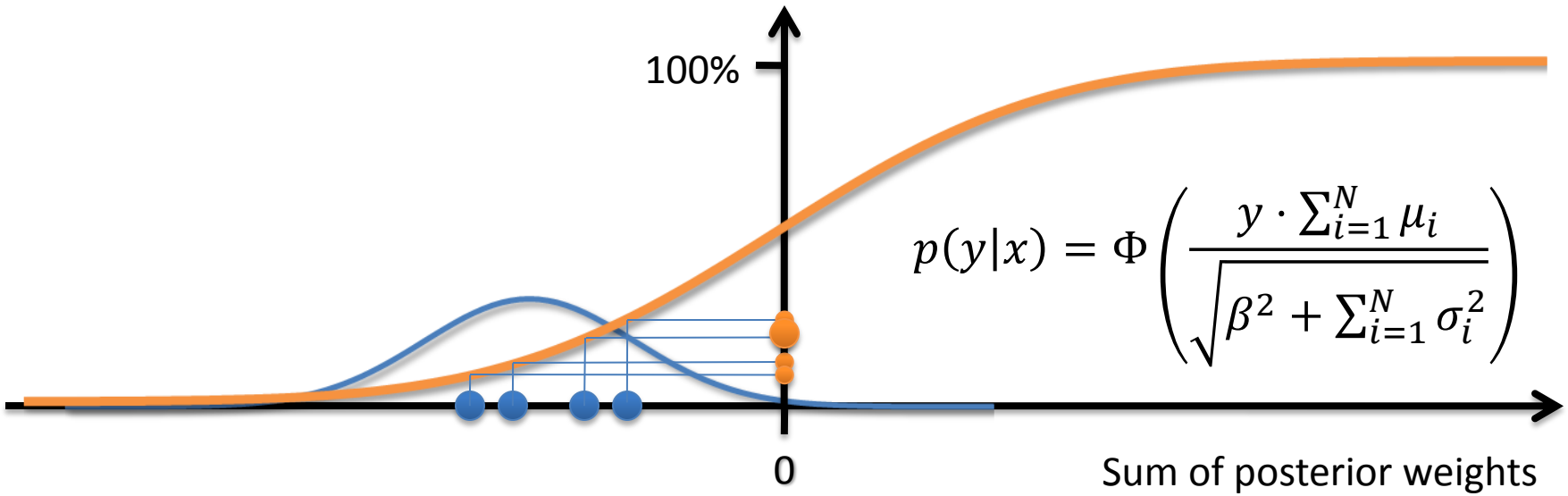
- The posterior to remain a **factorized Gaussian**
- Incremental **online learning** rather than batch

This is how it is done

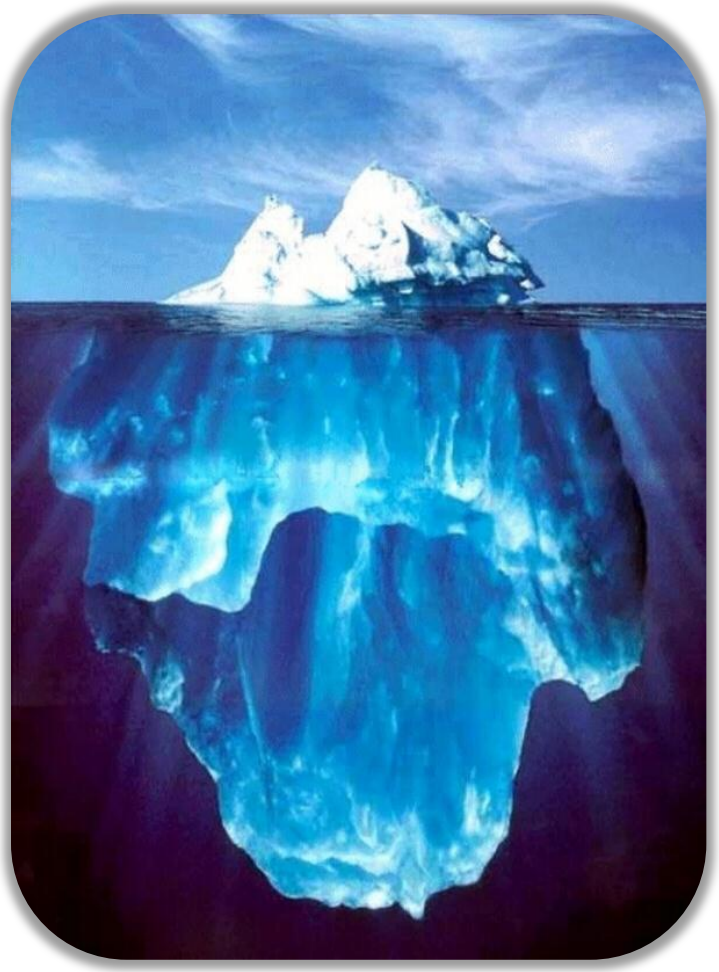
- Approximate inference with latent variables
- Single pass approximate (online) schedule

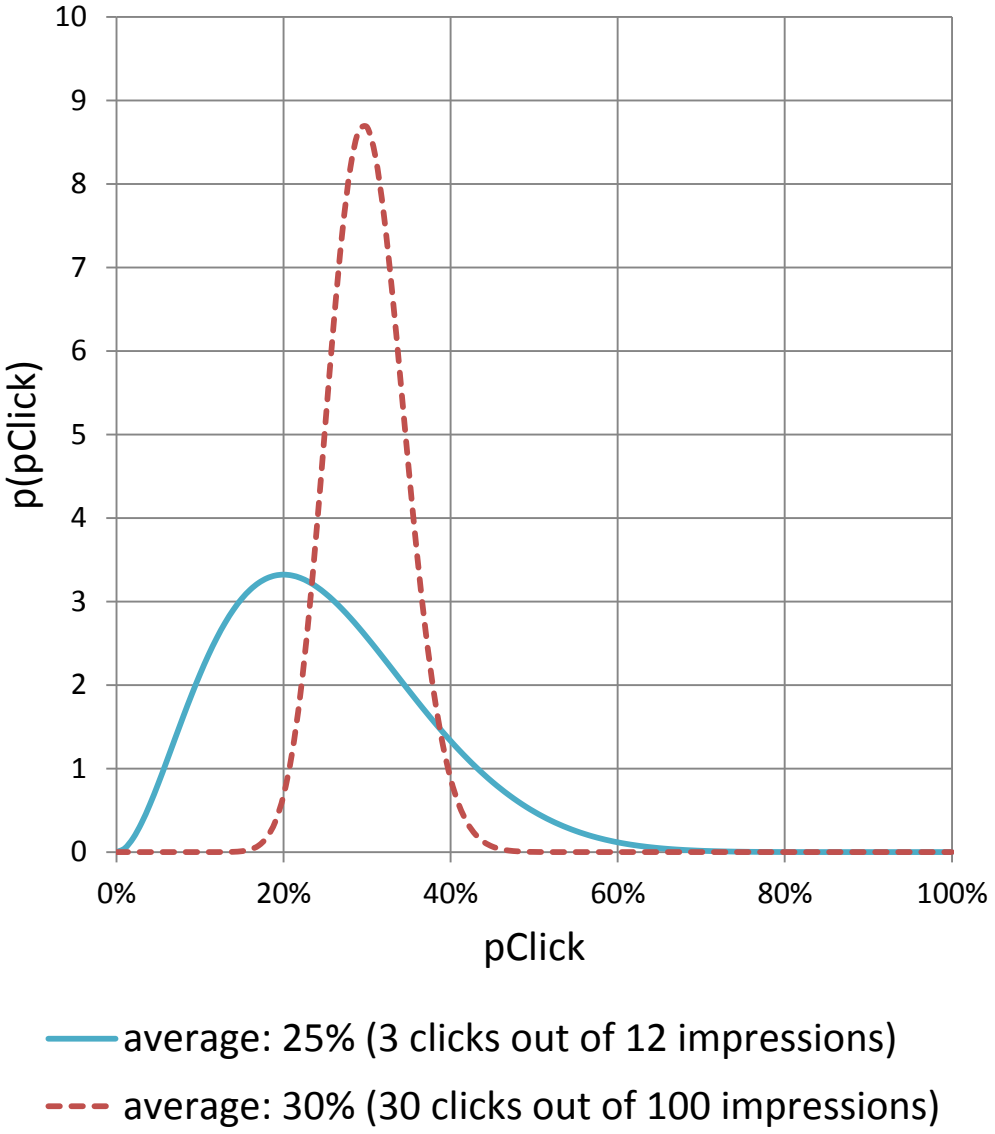
Predicting Average Probability of Click

Now that our posterior over the weights is a factorizing Gaussian…

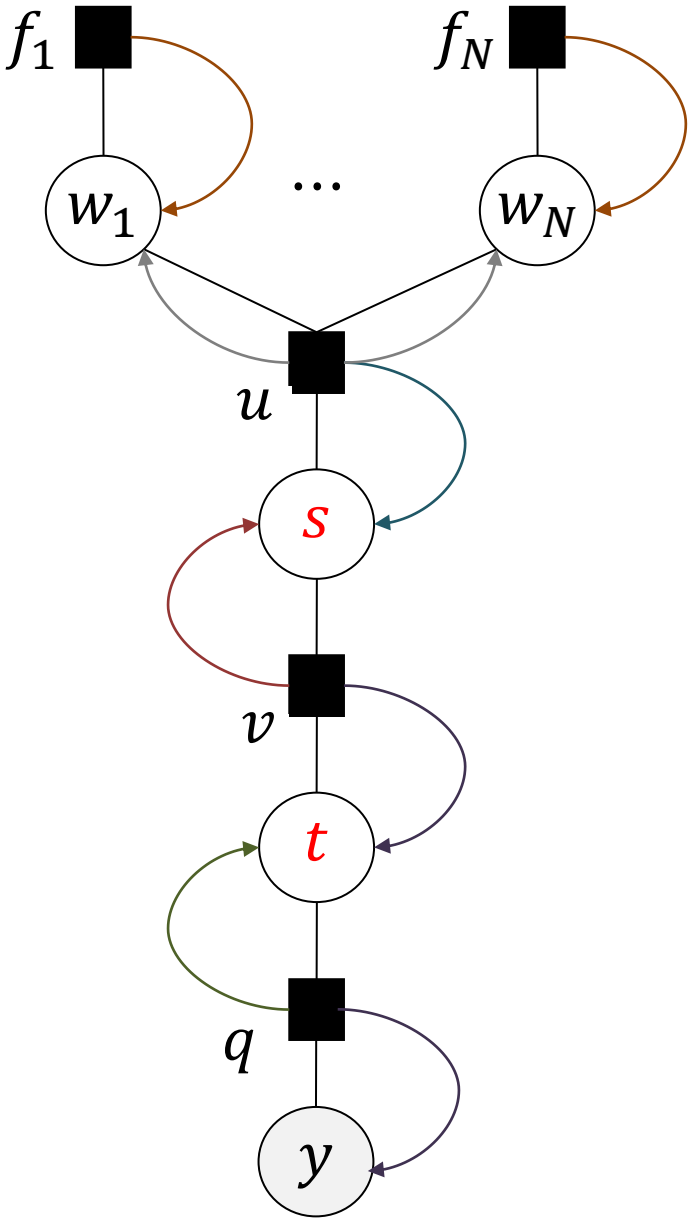


Principled Exploration



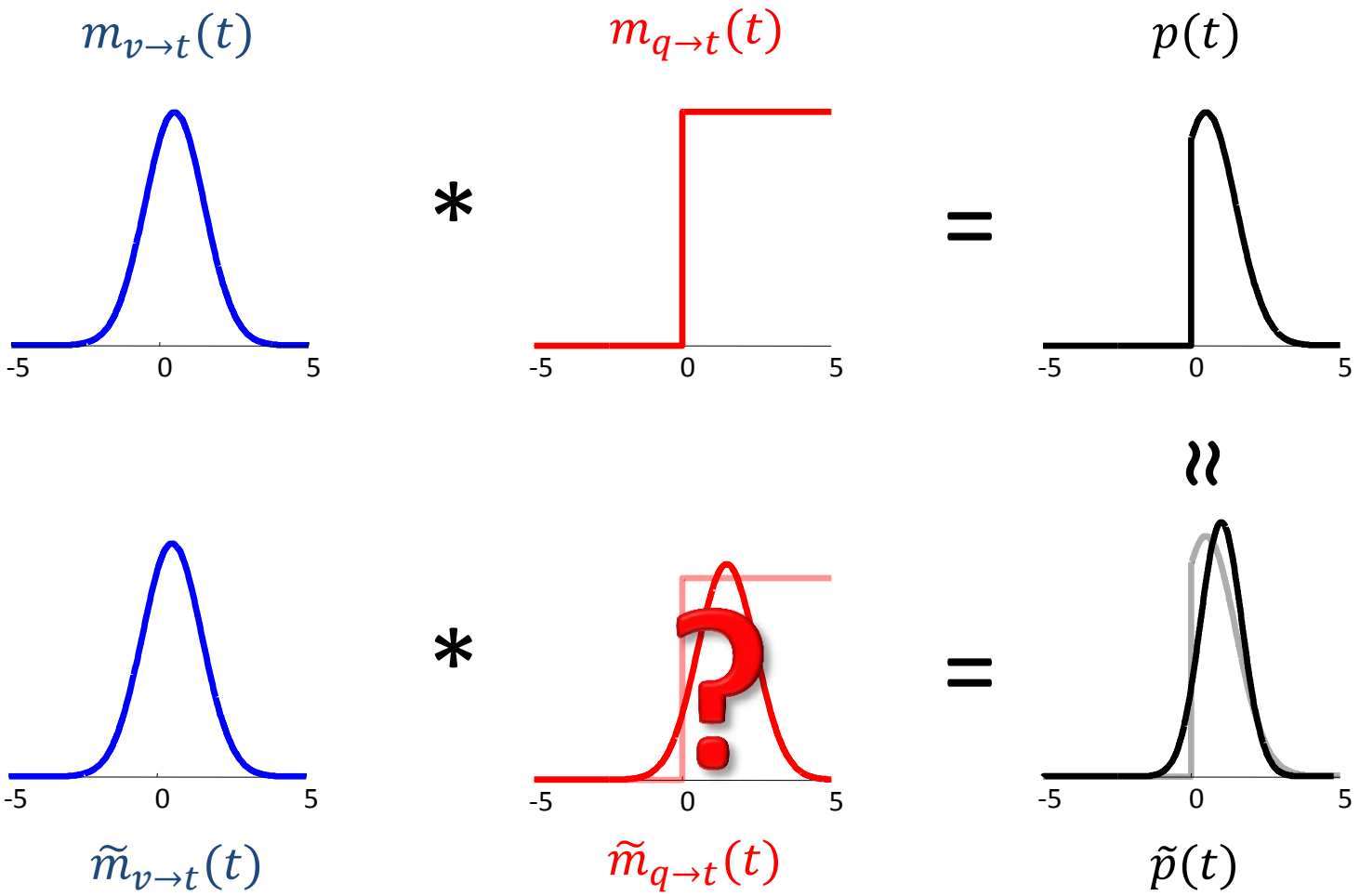


Approximate Inference with Latent Variables

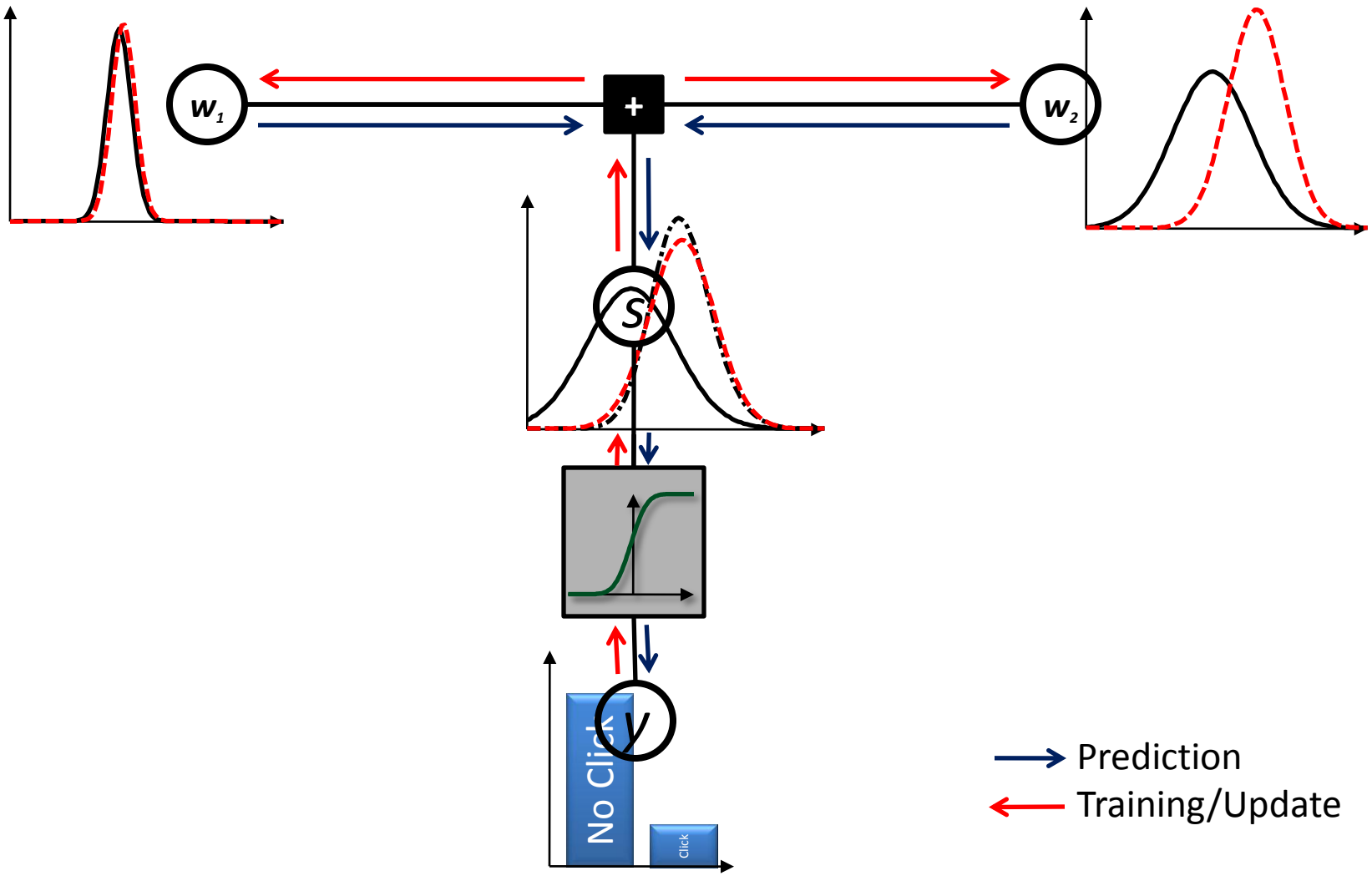


- Prior: $f_i(w_i) = \mathcal{N}\left(w_i; \mu_i, \sigma_i^2\right)$
- Sum of active weights: $u(s, \{w_i\}) = \delta\left(s - \sum_{i=1}^{N} w_i\right)$
- Noisy version thereof: $v(s, t) = \mathcal{N}(t; s, \beta^2)$
- The sign of t determines click: $q(t, y) = \delta(y - \text{sign}(t))$

Approximating $p(t)$ and $m_{q\to t}(t)$



Updating the Posterior



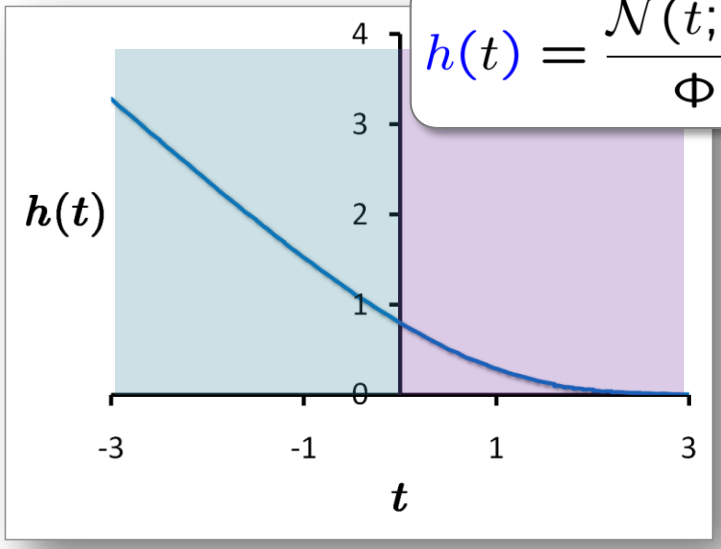
Posterior Updates for the Click Event

$$\mu_i \leftarrow \mu_i + \frac{\sigma_i^2}{s} \cdot h\left[\frac{\sum_{j=1}^{d} \mu_j}{s}\right]$$

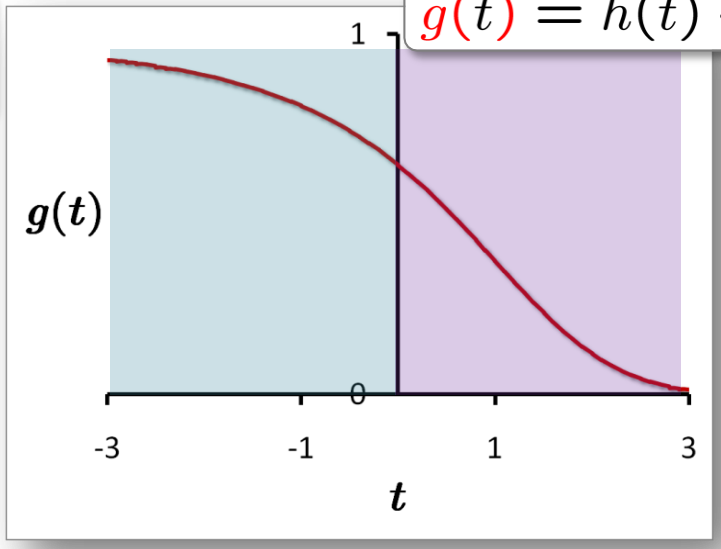
$$\sigma_i^2 \leftarrow \sigma_i^2 \left(1 - \frac{\sigma_i^2}{s^2} \cdot g\left[\frac{\sum_{j=1}^{d} \mu_j}{s}\right]\right)$$

$$s^2 = \beta^2 + \sum_{j=1}^{d} \sigma_j^2$$

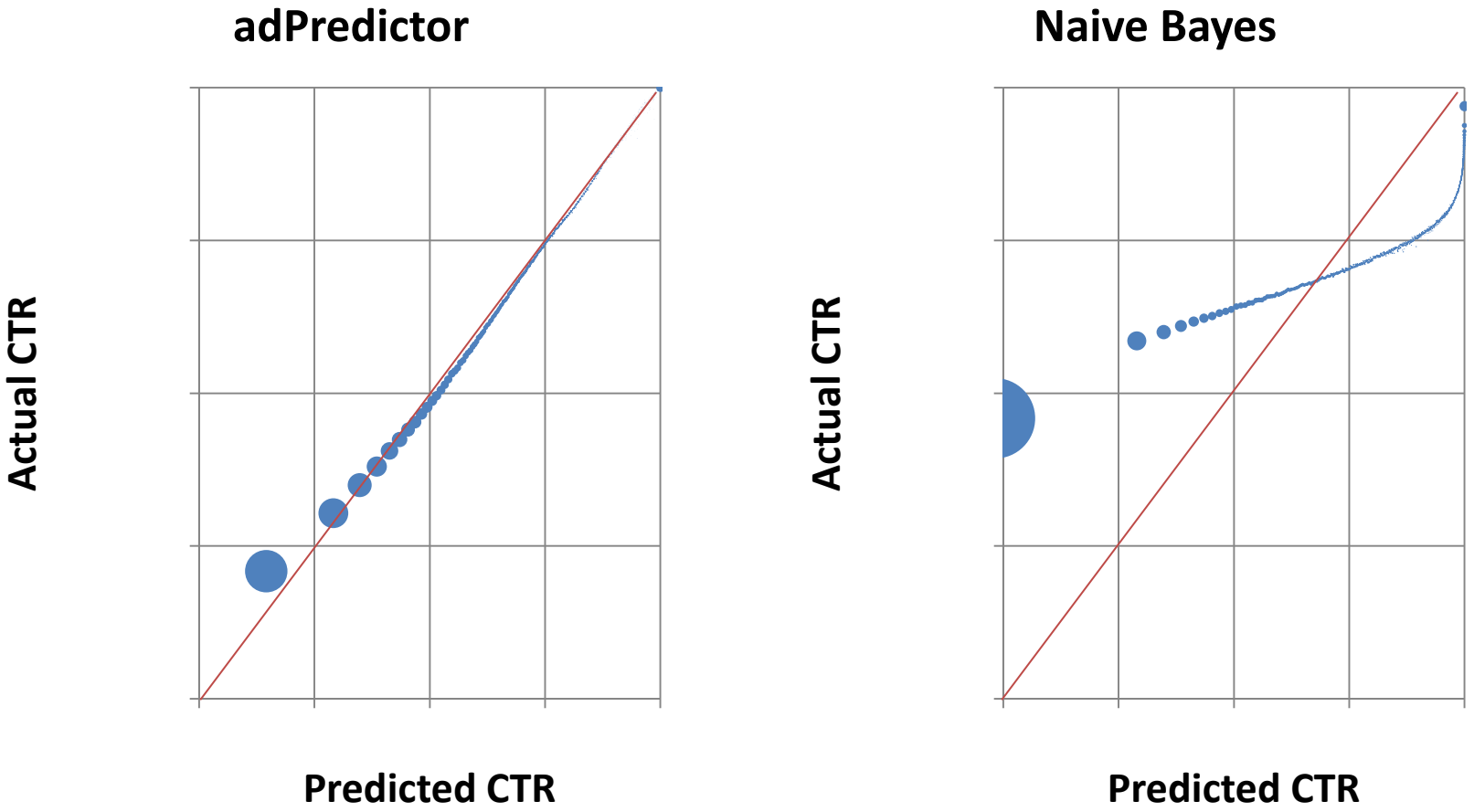
$$h(t) = \frac{\mathcal{N}(t;\, 0, 1)}{\Phi(t)}$$



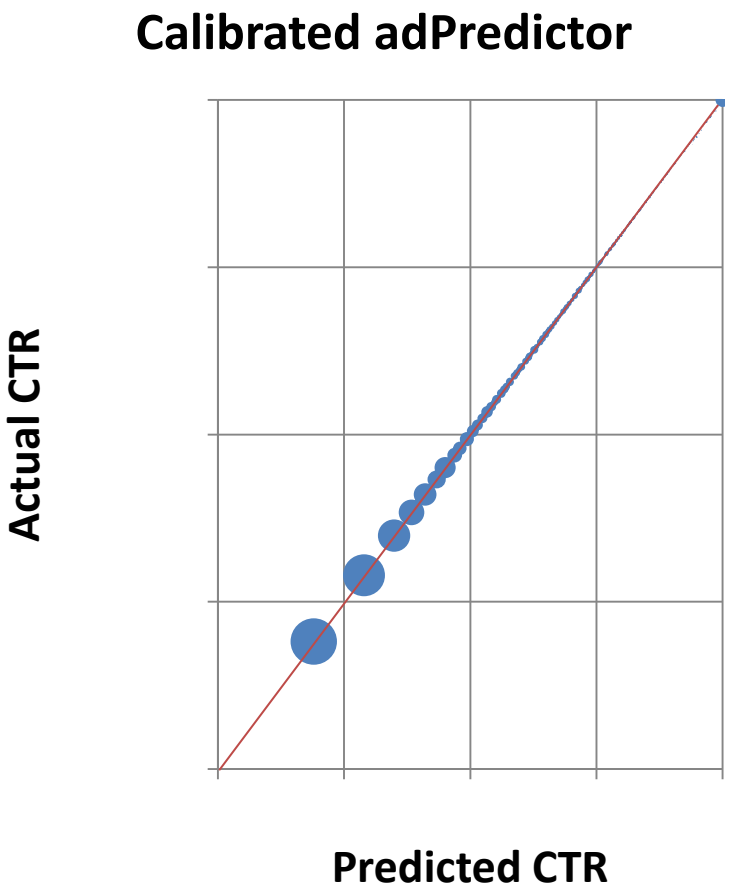
$$g(t) = h(t) \cdot [h(t) + t]$$

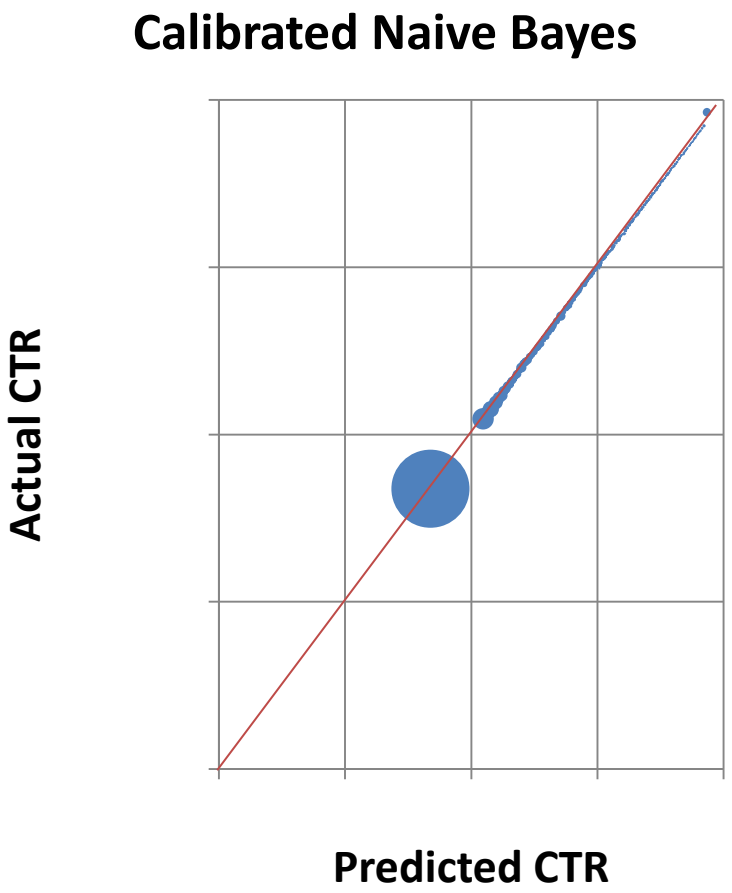


The importance of joint updates

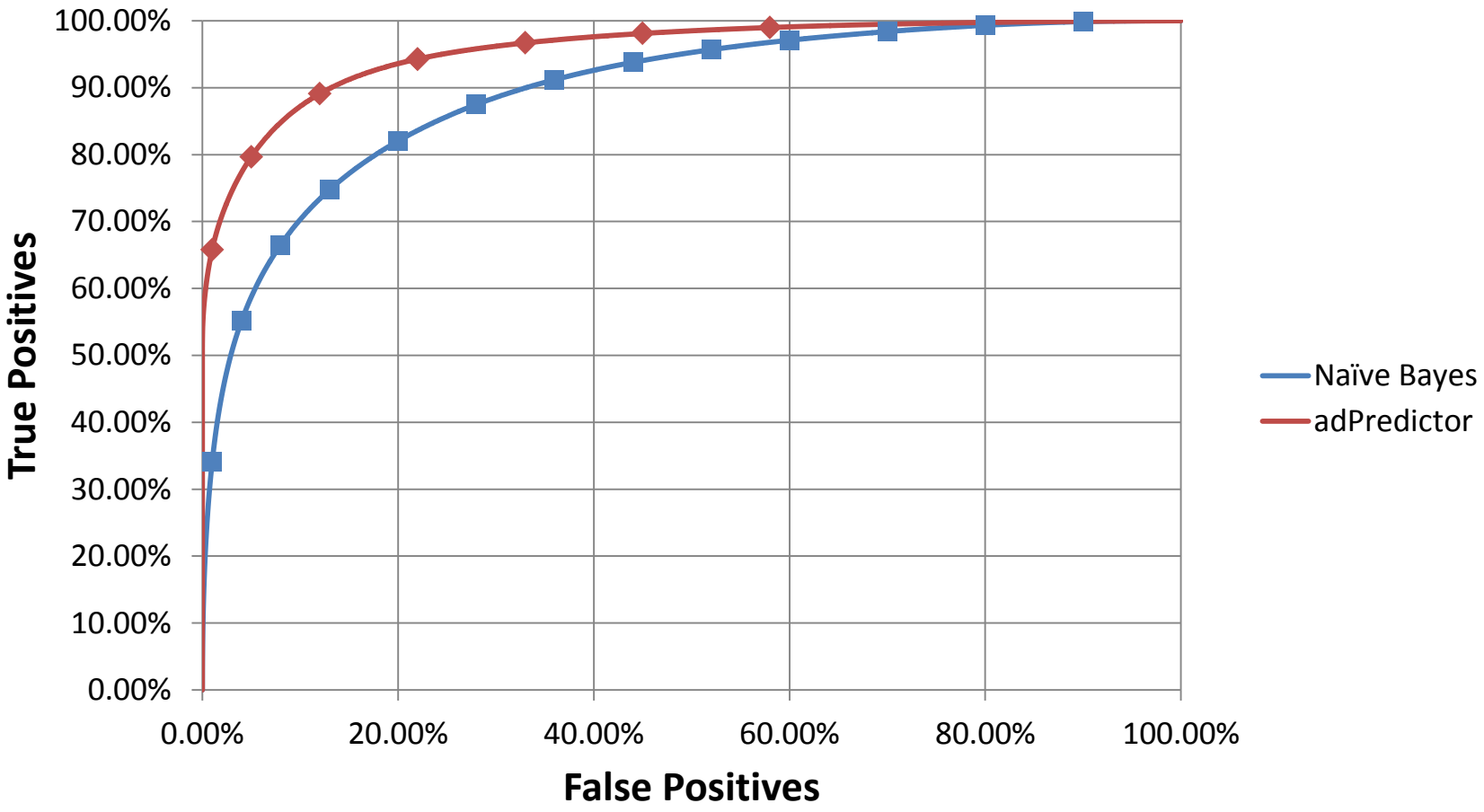


Calibration by Isotonic Regression





Calibration Can't Improve the ROC



adPredictor Wrap Up

- Automatic learning rate
- Calibrated: 2% prediction means 2% clicks
- Use of very many features, even if correlated
- Modelling the uncertainty explicitly
- Natural exploration mode

Discussion (For Later)

- Sample selection bias and exploration
- Dynamics: forgetting with time
- Pruning uninformative weights
- Approximate parallel inference
- Hierarchical priors
- Input features… the secret sauce

Some of this is detailed in the ICML 2010 paper:

Web-Scale Bayesian Click-Through Rate Prediction for Sponsored Search Advertising in Microsoft's Bing Search Engine

We are hiring! Please contact me if you are interested.

Thank you!

joaquinc@microsoft.com

We are hiring! Please contact me if you are interested.