

Simulation results

Sampling with replacement

I've enabled sampling with replacement for the optimal and random teachers. It doesn't seem fair to enforce sampling with replacement for the grid teacher, since the grid teacher covers a set of evenly spaced points by definition. With a linear boundary in 2D and a linear SVM user model, the optimal teacher seems to pick 3 distinct examples and then repeat the third one forever (Figure 1). As Anca said, this is a useful way to determine that the optimal teacher has converged.

Additional user models

I've added basically all of the user models Sandy had:

- SVM with RBF kernel
 - $C = 1, \gamma = 0.1$ (pretty default parameters)
 - $C = 528, \gamma = 0.001$ (the best parameters from Sandy's grid search for fitting to real user data)
- Online kernel machine (OKM) as in Anca's RoMan paper (2012). This performs function approximation of $P(Y = 1|X)$ by adding a Gaussian bump (positive or negative) for every incoming example. The contributions of old examples are reduced over time.
 - $\eta = 0.85, \lambda = 0.05, w = 1$ (These draw upon Sandy's parameters from grid search, though I may have picked the wrong ones.) Starts with an uninformative "prior"/initial guess of $P(Y = 1|X) = 0.5$.
- KDE
 - Gaussian kernel with bandwidth $= I_d$ by default
- Generalized context model (GCM), a model from cognitive science where the class-conditional density (unnormalized) of x depends on class examples x_1^n as follows: $s(x, x_1^n) = \sum_{i=1}^n \exp(-cd(x, x_i)^\alpha)$. Reducing kernel is not yet implemented.
 - $c = 1, \alpha = 2, d(x, y) = \frac{1}{d} \|x - y\|_r, r = 1$. These are pretty default parameters, with $r = 1$ chosen to differentiate GCM more from KDE.

The first two are discriminative; the latter two are generative.

We can easily run any of these user models with other parameter settings. I haven't really put time into tuning parameter settings. This regime where the teacher has complete knowledge of the user model is overly simplified anyway.

Nonlinear boundaries

To go along with the additional user models, I tried a variety of nonlinear boundaries:

- Polynomial of degree m that passes through $m + 1$ randomly sampled points in the grid. I tried $m = 2, 3, 4$; you can set it as high as you like.
- Hardcoded functions like e^x , $\sin x$ and $x \sin x$

For sample boundaries, see Figure 2, 3, 4.

Observations

Random teacher degrades in high dimensions again

As we saw for teaching a linear boundary to a linear SVM, randomly teaching a nonlinear boundary to an RBF SVM degrades in higher dimensions. Compare the accuracy of randomly teaching a degree-2 polynomial

Active learning with linear SVM user model, optimal teacher
 13x6 grid with linear boundary ($w=[-0.62 \ 0.24]$, $b=-3.00$)

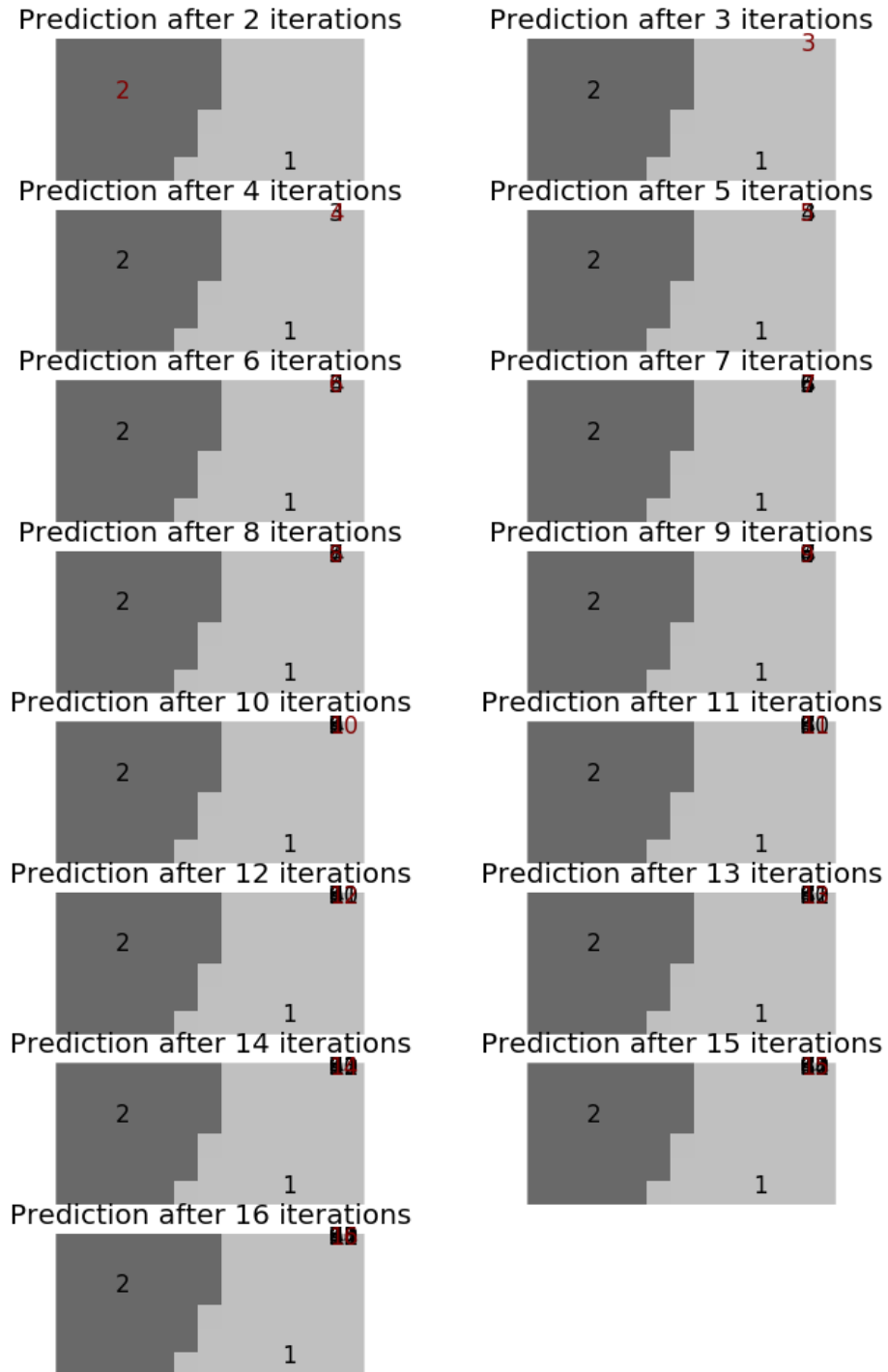


Figure 1: Linear user model, linear SVM, optimal teacher

Ground truth: 13x6 grid with
polynomial boundary of degree 2
passing through (3,4), (6,2), (8,2)
(w=[-8.4 1.87 -0.13 1.])



Figure 2: quadratic boundary

Ground truth: 13x6 grid with
polynomial boundary of degree 3
passing through (0,2), (4,4), (6,0), (8,4)
(w=[-2. -6.42 2.19 -0.18 1.])



Figure 3: cubic boundary

Ground truth: 13x6 grid with
boundary $2 * \sin(x)$



Figure 4: sinusoidal boundary

in x, y with free z coordinate (Figure 5) to the accuracy of randomly teaching a degree-2 polynomial in just x, y (Figure 6).

Grid teaching can get worse with more examples

A grid teacher can be poor at teaching nonlinearities to an RBF SVM, since examples are predictable and sparse. The 95th percentile grid teacher sees accuracy decline slightly as more examples come in. This happens on a quadratic boundary (Figure 6), higher-degree polynomial boundaries, and sinusoidal boundaries (Figure 7).

With the right parameters, RBF SVM can be taught nonlinear boundaries

An RBF SVM with parameter settings $C = 1, \gamma = 0.1$ is taught by an optimal teacher to a higher final accuracy than an RBF SVM with $C = 528, \gamma = 0.001$ on “wiggly” boundaries like degree 3 and 4 polynomials and sinusoidal functions. This is predictable, since a smaller γ means that examples farther away exert more influence over the learned boundary. Larger C places more weight on the loss, preferring a more complex boundary, but the smaller γ seems more influential here. Compare the accuracy of optimally teaching an RBF SVM a cubic boundary in Figures 8 ($C = 1, \gamma = 0.1$) and 9 ($C = 528, \gamma = 0.001$).

With more complex boundaries, the optimal teacher starts repeating examples later. For instance, teaching a quadratic boundary to an RBF SVM, all examples after the 7th are repeats (Figure 10).

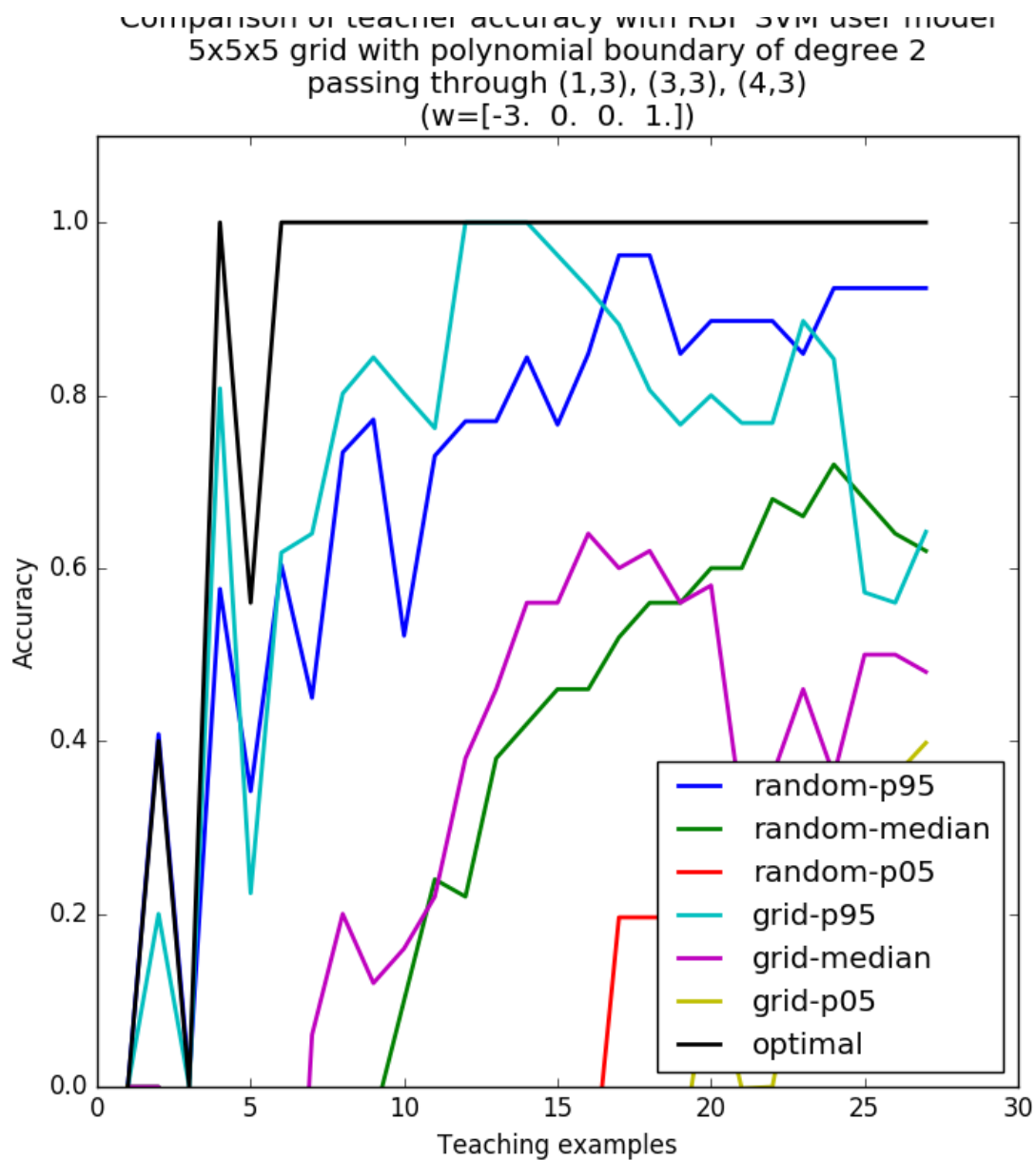


Figure 5: 5x5x5, quadratic boundary, RBF SVM user model

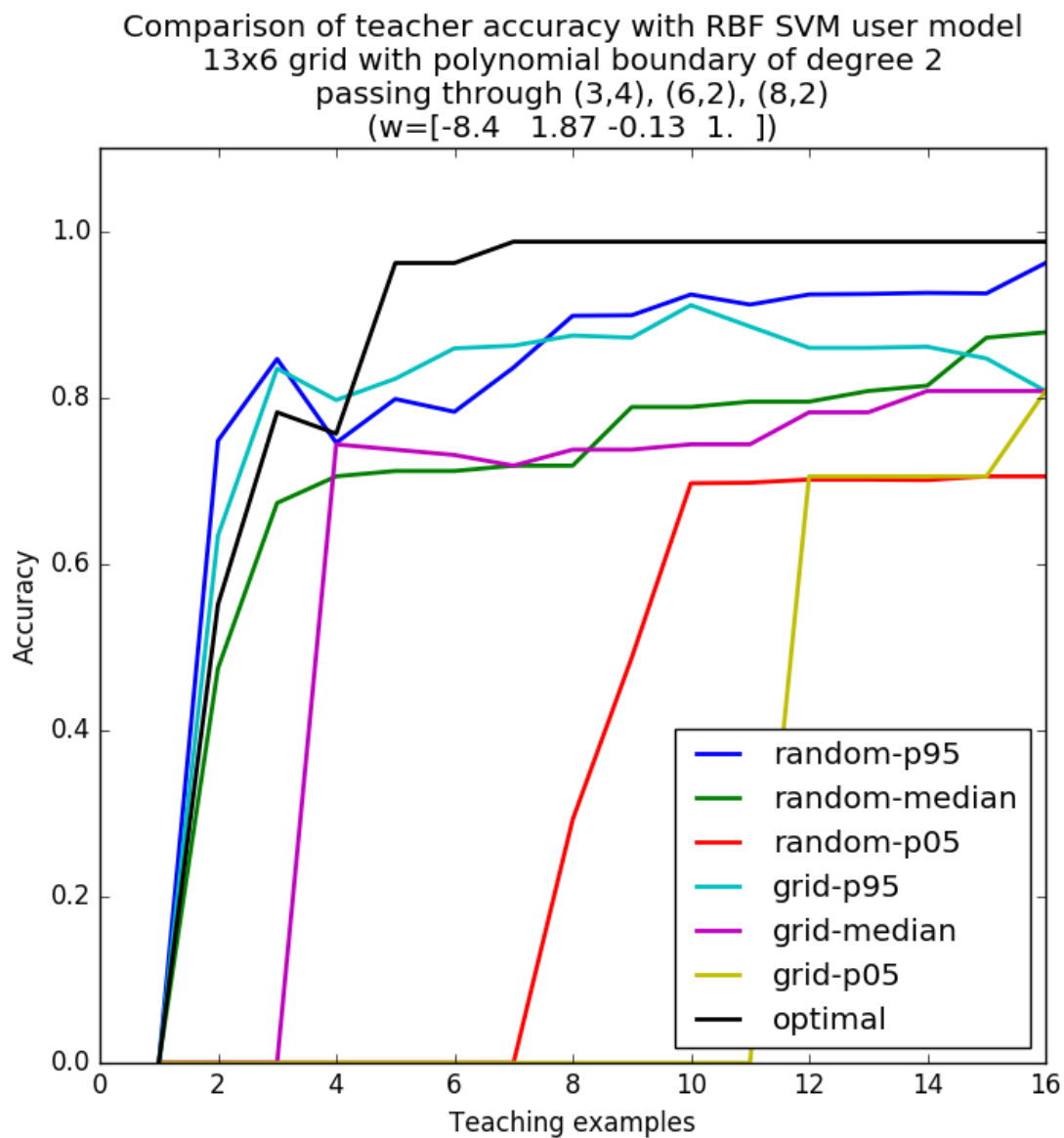


Figure 6: 13x6, quadratic boundary, RBF SVM user model

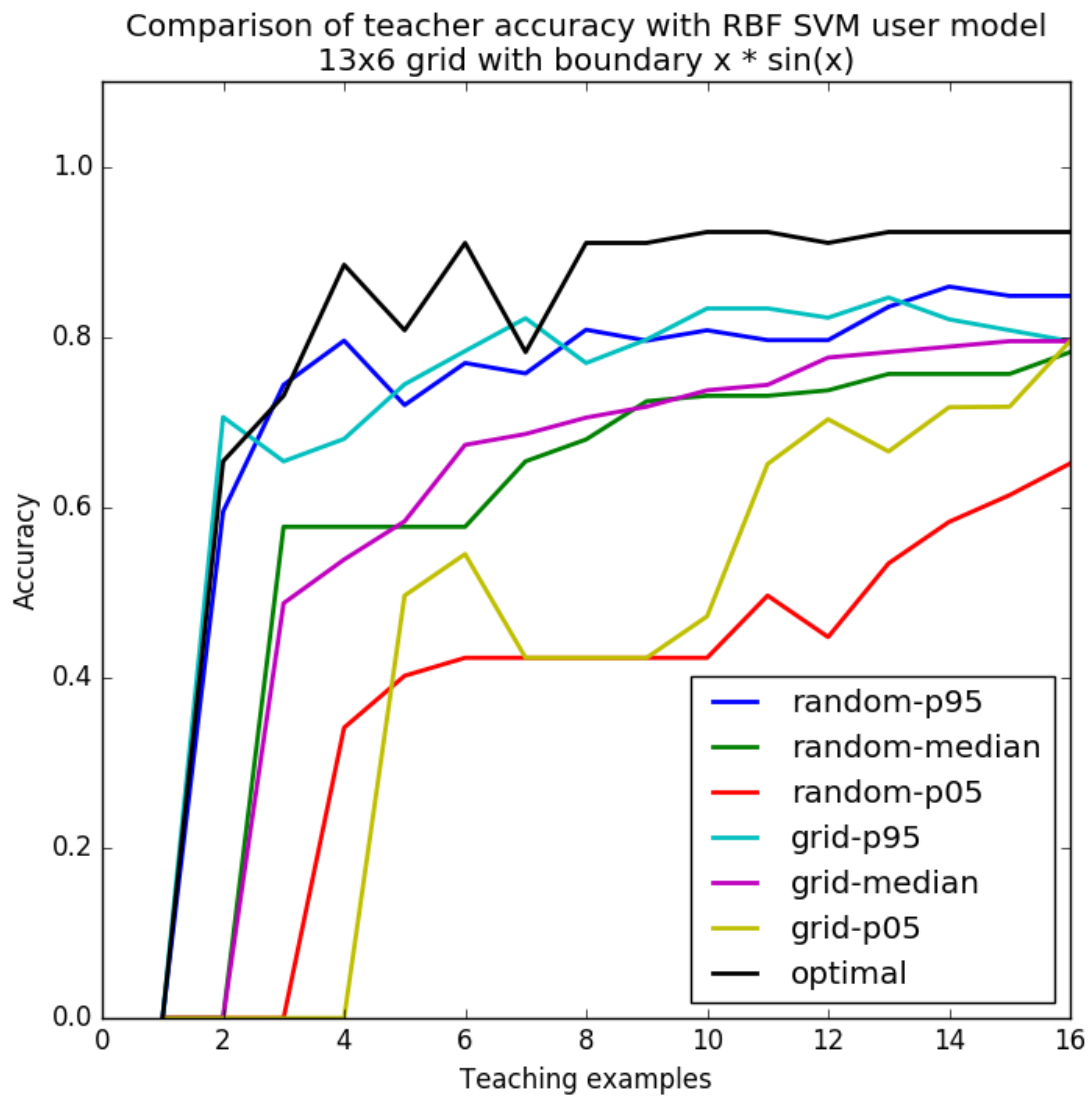


Figure 7: 13x6, $x \sin x$ boundary, RBF SVM user model

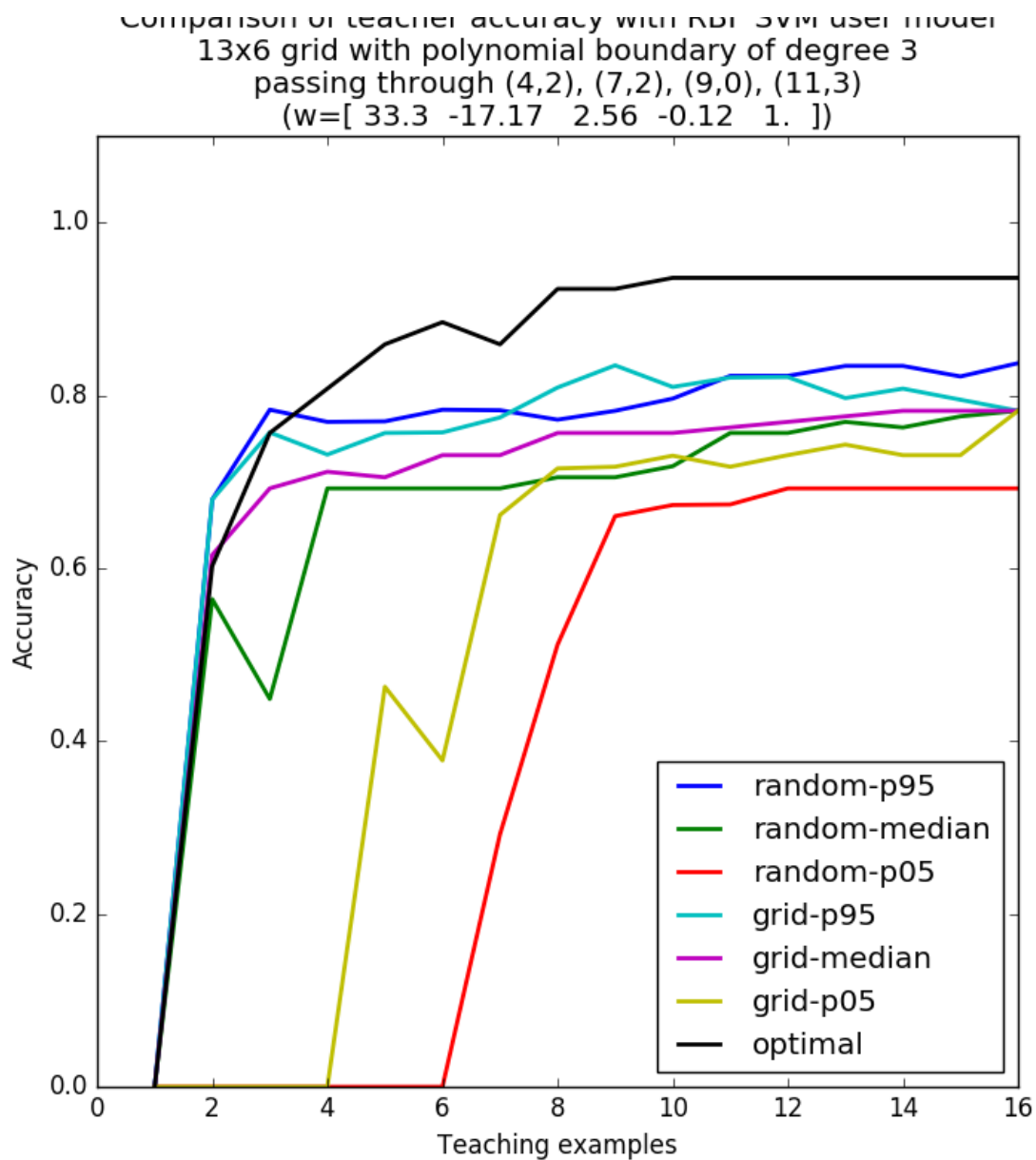


Figure 8: 13x6, cubic boundary, RBF SVM user model, $C=1$, $\gamma=0.1$

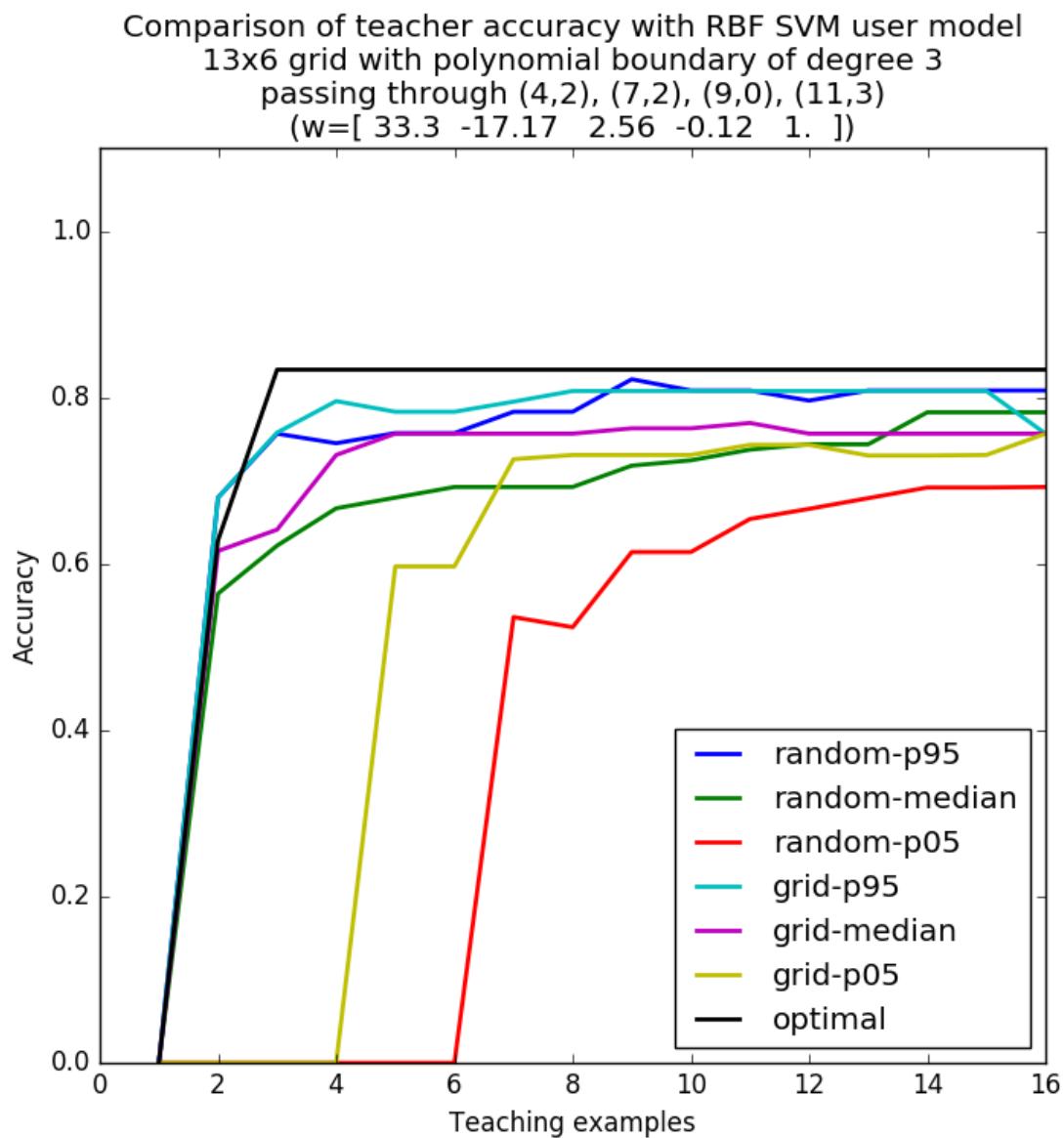


Figure 9: 13x6, cubic boundary, RBF SVM user model, $C=528$, $\gamma=0.001$

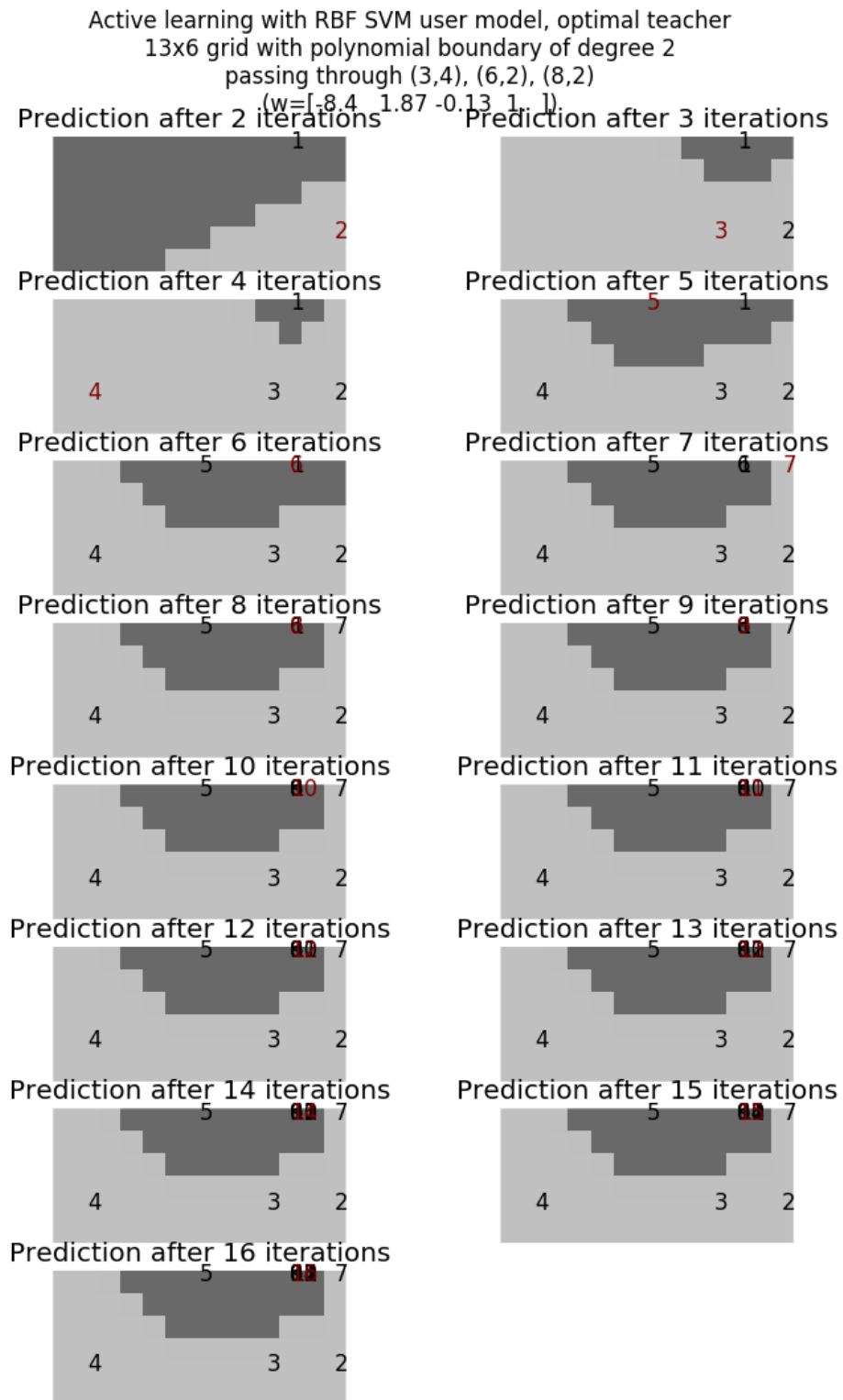


Figure 10: 13x6, quadratic boundary, RBF SVM user model, optimal teaching

Optimal teaching creates greater improvement on nonlinear boundaries when the user model is KDE or OKM

Optimally teaching an RBF SVM a complex boundary like cubic polynomials and sinusoids hits a ceiling below 100% accuracy. KDE and OKM seem like more flexible user models at the given parameter settings, since they can be optimally taught to close to 100% accuracy. To see this happen on the $\sin x$ boundary, consult Figures 11, 12, 13.

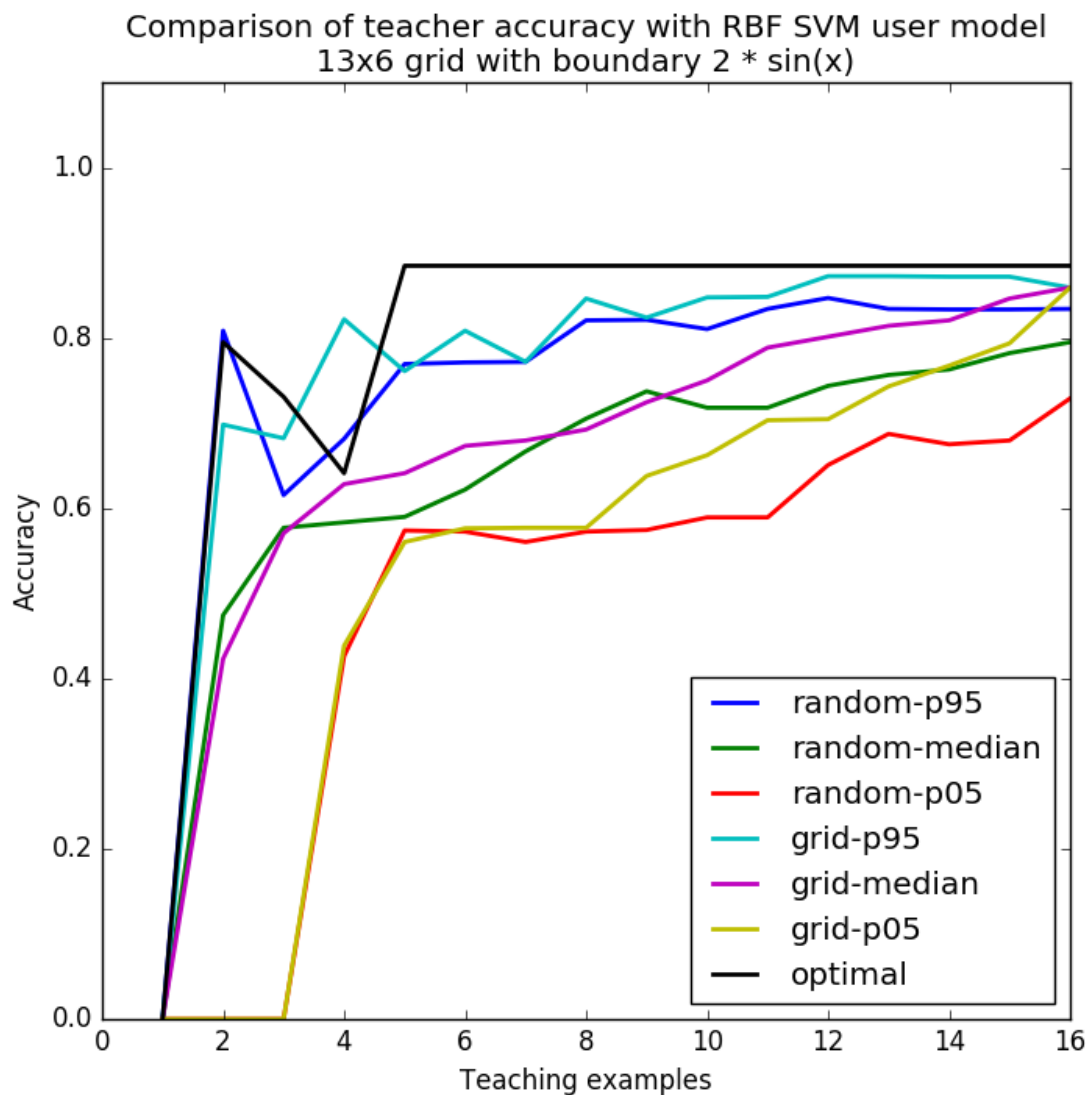


Figure 11: 13x6, $\sin x$ boundary, RBF SVM user model

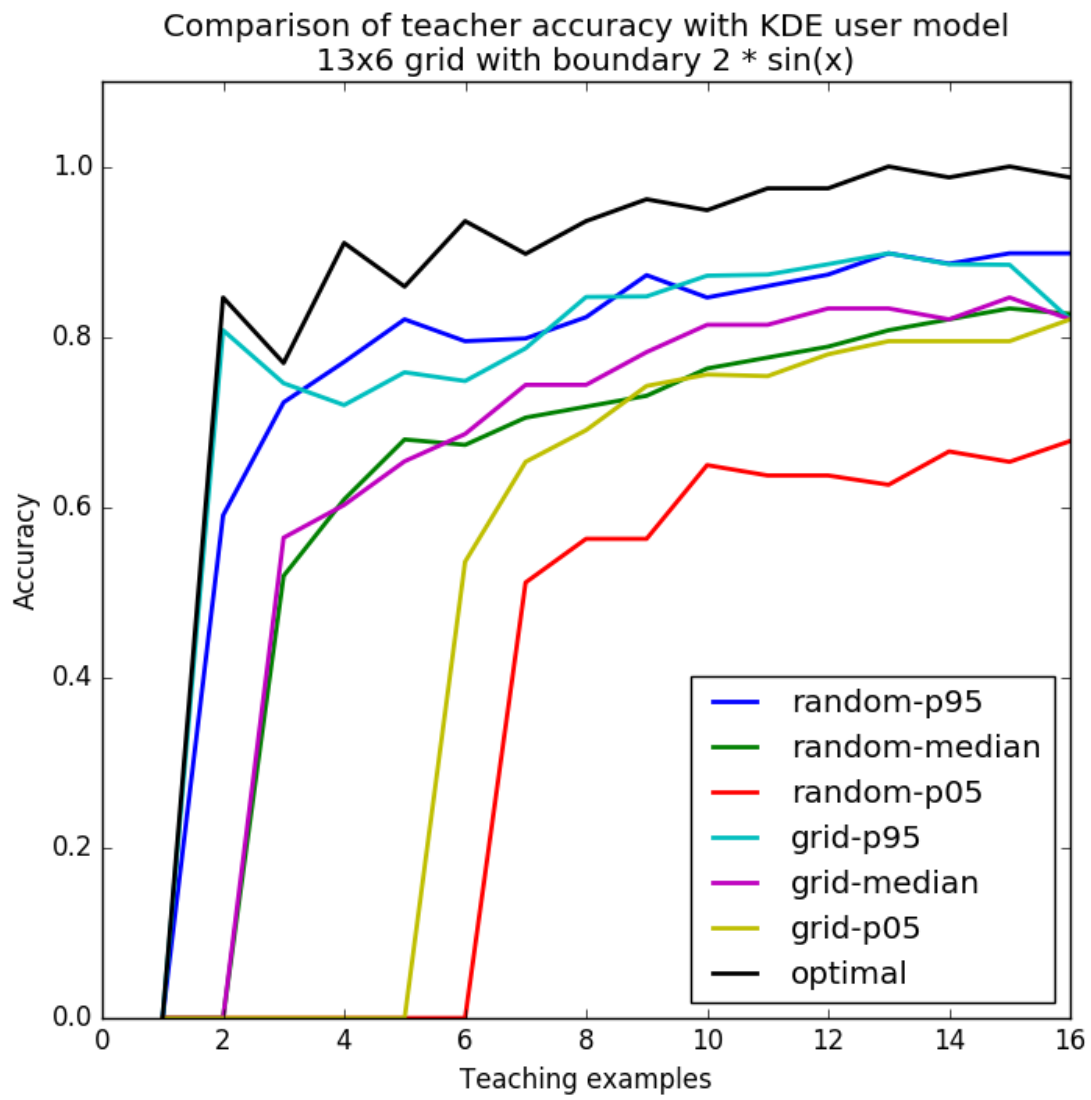


Figure 12: 13x6, $\sin x$ boundary, KDE user model

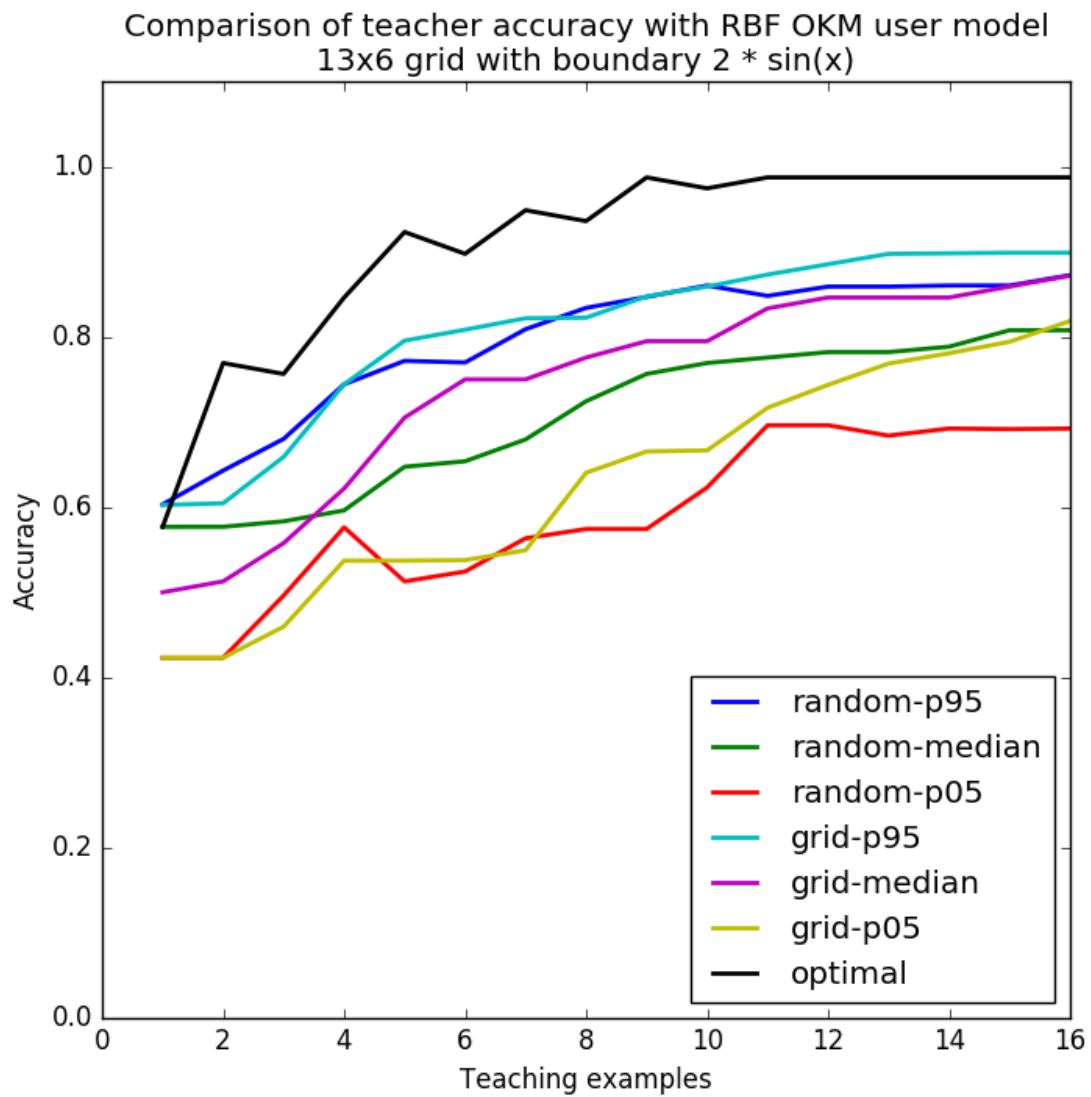


Figure 13: 13x6, $\sin x$ boundary, OKM user model

Optimally teaching an OKM results in high uncertainty across the grid

Predictions from user models are obtained by thresholding $P(Y = 1|X)$ at 0.5; the optimal teacher seeks to minimize prediction error. The actual values of $P(Y = 1|X)$ could be close to 0.5, however. Prediction certainty is not currently factored into our notion of optimality. As a result, optimal teaching of an OKM model results in uncertainty across large parts of the grid, whether the boundary is linear (Figure 14), higher-degree polynomial or exponential. By contrast, KDE at its default parameter setting allows for more certain predictions (Figure 15).

Active learning with RBF OKM user model, optimal teacher
 13x6 grid with linear boundary ($w=[-0.62 \ 0.24]$, $b=-3.00$)

Evaluation after 1 iterations



Evaluation after 2 iterations



Evaluation after 3 iterations



Evaluation after 4 iterations



Evaluation after 5 iterations



Evaluation after 6 iterations



Evaluation after 7 iterations



Evaluation after 8 iterations



Evaluation after 9 iterations



Evaluation after 10 iterations



Evaluation after 11 iterations



Evaluation after 12 iterations



Evaluation after 13 iterations



Evaluation after 14 iterations



Evaluation after 15 iterations



Evaluation after 16 iterations



Figure 14: 13x6, linear boundary, OKM user model

Active learning with KDE user model, optimal teacher
13x6 grid with linear boundary ($w=[-0.62 \ 0.24]$, $b=-3.00$)

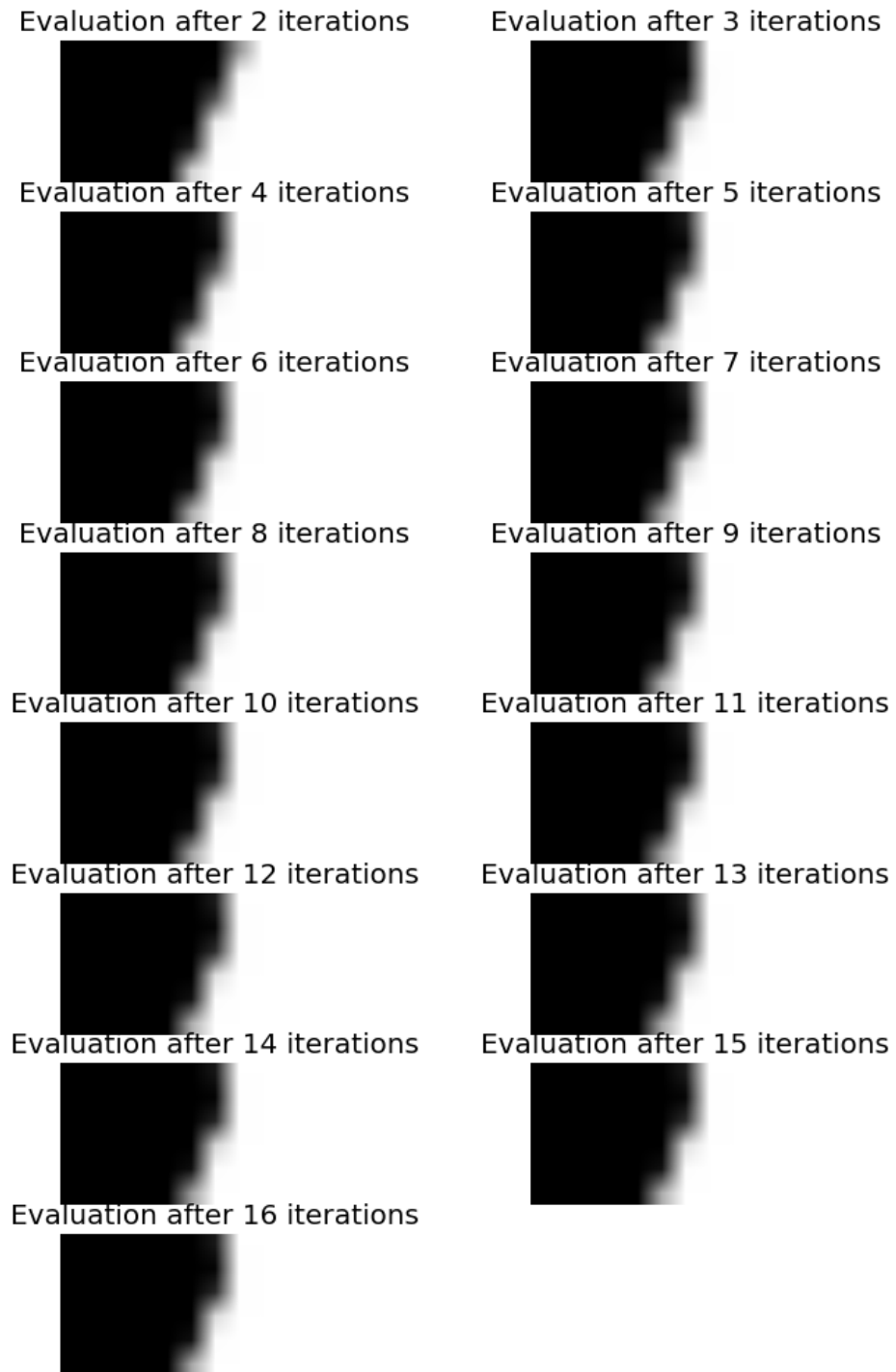


Figure 15: 13x6, linear boundary, KDE user model