Deep Probability Estimation

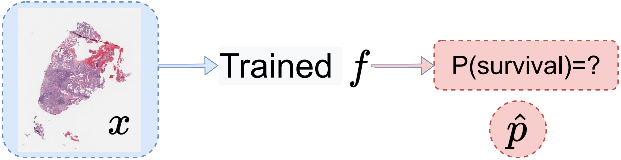
Sheng Liu*, Aakash Kaku*, Weicheng Zhu*, Matan Leibovich*, Sreyas Mohan*, Boyang Yu, Haoxiang Huang, Laure Zanna, Narges Razavian, Jonathan Niles-Weed, Carlos Fernandez-Granda

Probability Estimation

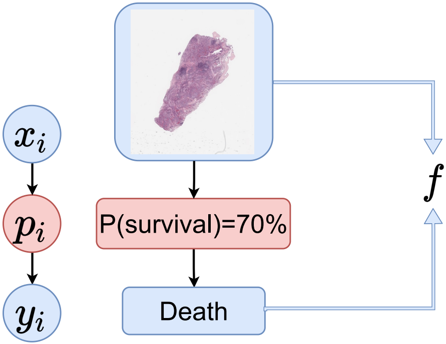
Goal: Estimate probability of uncertain events from high-dimensional input (images, videos)

Not equivalent to classification because of inherent (aleatoric) uncertainty

Probability estimation via deep learning



Training



Synthetic Dataset: Face-based Risk Prediction

x_i

Face-Age dataset from *Age Progression/Regression by Conditional Adversarial Autoencoder*. Zhang, Z., Song, Y., and Qi, H., CVPR 2017

Synthetic Dataset: Face-based Risk Prediction

x_i	Age z_i
	20

Face-Age dataset from *Age Progression/Regression by Conditional Adversarial Autoencoder*. Zhang, Z., Song, Y., and Qi, H., CVPR 2017

Synthetic Dataset: Face-based Risk Prediction

x_i	Age z_i	$p_i = \frac{z_i}{100}$
	20	0.2

Face-Age dataset from *Age Progression/Regression by Conditional Adversarial Autoencoder*. Zhang, Z., Song, Y., and Qi, H., CVPR 2017

Synthetic Dataset: Face-based Risk Prediction

x_i	Age z_i	$p_i = \frac{z_i}{100}$	y_i
	20	0.2	0

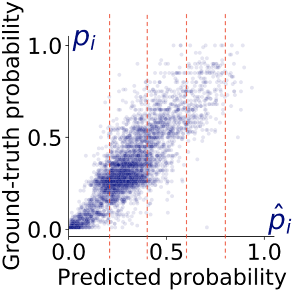
Face-Age dataset from *Age Progression/Regression by Conditional Adversarial Autoencoder*. Zhang, Z., Song, Y., and Qi, H., CVPR 2017

Synthetic Dataset: Face-based Risk Prediction

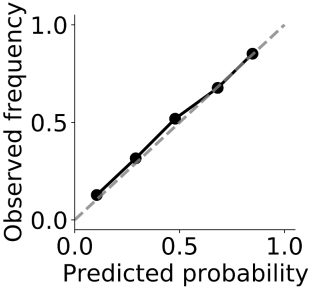
x_i	Age z_i	$p_i = \frac{z_i}{100}$	y_i
	20	0.2	0
	70	0.7	1

Face-Age dataset from *Age Progression/Regression by Conditional Adversarial Autoencoder*. Zhang, Z., Song, Y., and Qi, H., CVPR 2017

Evaluation



Ground-truth probability
p_i
1.0
0.5
0.0
$\hat{p}_i$
0.0
0.5
1.0
Predicted probability



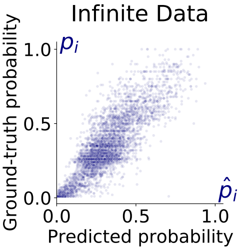
Observed frequency
1.0
0.5
0.0
0.0
0.5
1.0
Predicted probability

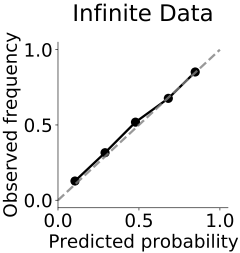
Cross-entropy

Standard cross-entropy loss is a *proper scoring rule*

Probabilities estimated by minimizing cross entropy are well calibrated in an "infinite" data regime

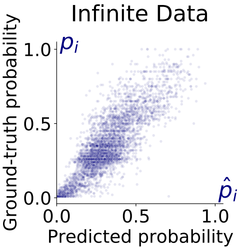
Face-based Risk Prediction

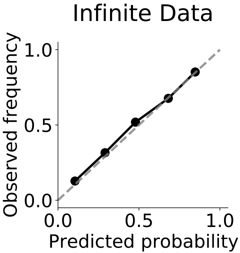




Face-based Risk Prediction

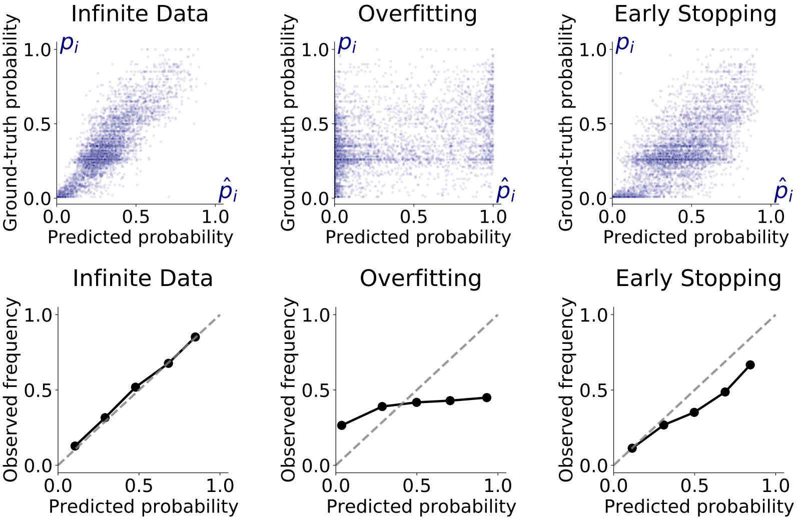
What happens if dataset is finite?





Face-based Risk Prediction

What happens if dataset is finite?

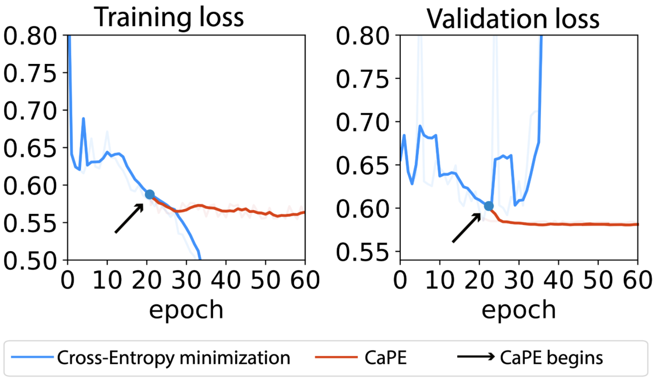


Calibrated Probability Estimation (CaPE)

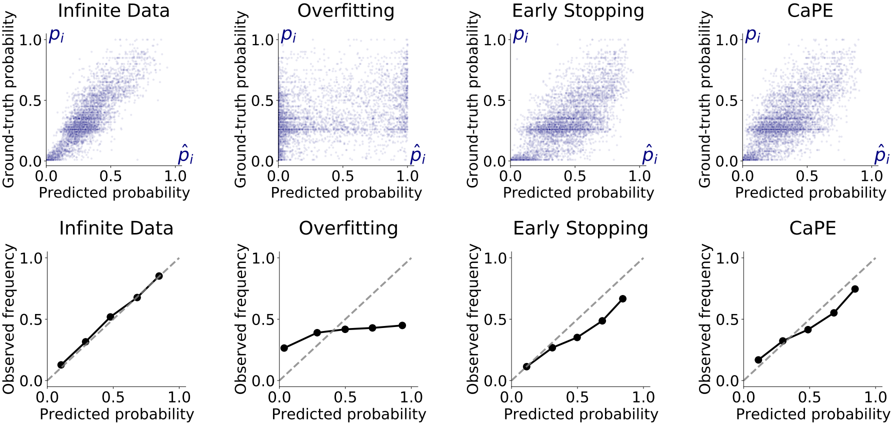
Goal: Improve model *while remaining calibrated*

1. Minimize cross-entropy loss until memorization begins
2. Alternate between:
   - Enforcing calibration
   - Minimizing cross entropy

Calibrated Probability Estimation (CaPE)



Face-based Risk Prediction

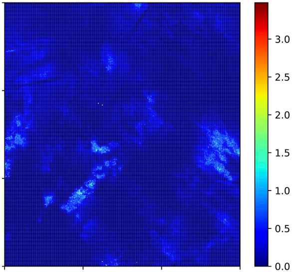


Real-world datasets

Weather Forecasting

Will it rain?

Radar map of past precipitation



Collision Prediction

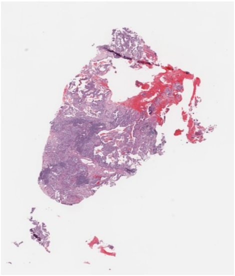
Will there be a collision?

Dashboard video

Survival Forecasting

Will a patient die in 5 years?

Histopathology image



Results

The Brier Scores of baseline models, scaled by 0.01.

Method	Cancer Survival	Weather forecasting	Collision Prediction
CE early-stop	23.96	20.57	8.59
Temperature	23.73	20.21	8.51
Platt Scaling	23.33	19.53	8.23
Dirichlet Cal.	24.08	21.89	8.78
Mix-n-match	23.67	20.21	8.52
Focal Loss	23.31	20.27	9.82
Entropy Reg.	23.62	19.77	11.10
MMCE Reg.	23.73	20.12	8.48
Deep Ens.	23.47	18.82	8.55
CaPE (bin)	23.20	**18.37**	8.18
CaPE (kern.)	**23.18**	18.39	**8.13**

Conclusions

- Deep networks can be effective for probability estimation

Conclusions

- Deep networks can be effective for probability estimation
- Exploiting early learning and enforcing calibration can improve probability estimates

Conclusions

- Deep networks can be effective for probability estimation
- Exploiting early learning and enforcing calibration can improve probability estimates
- More benchmark datasets are needed!

Acknowledgements

Research partially supported by NSF and NIH