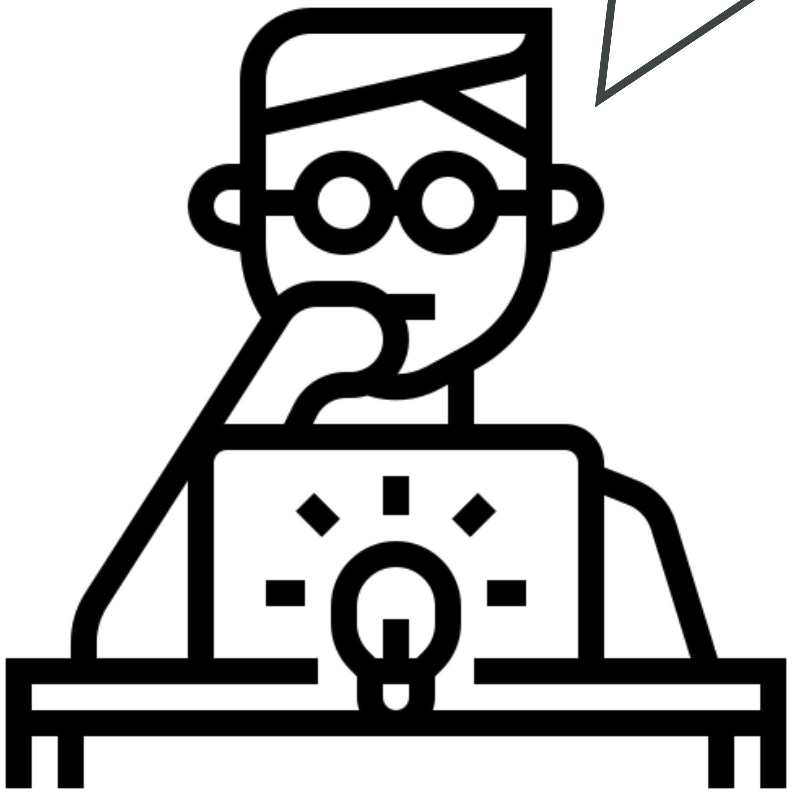
Understanding Model Uncertainty and Enhancing Probability Calibration in NN

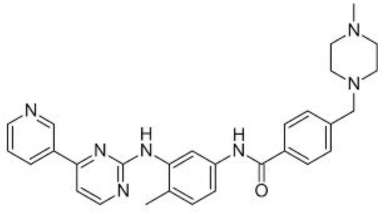
Understanding Model Uncertainty in NN

***Uncertainty Quantification* in NN?**
How? And why?

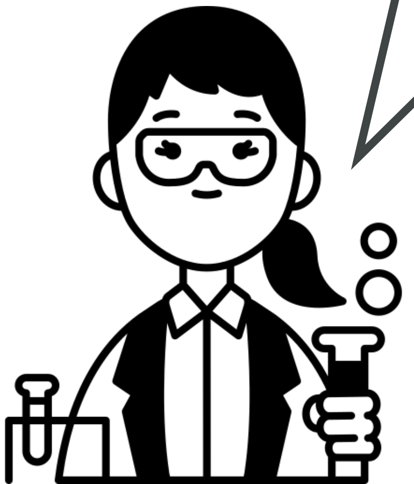


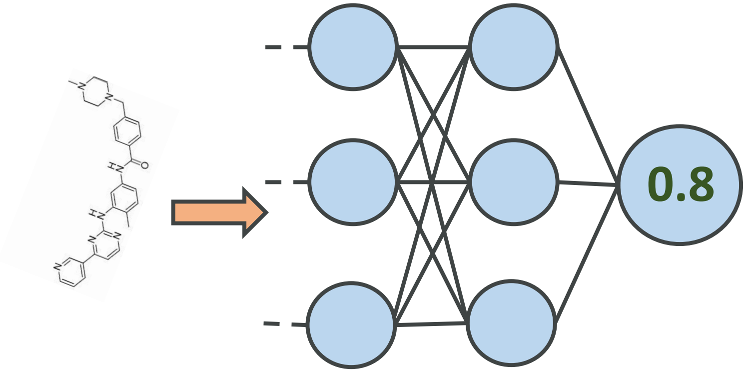
Why do we need Uncertainty Quantification (UQ)?

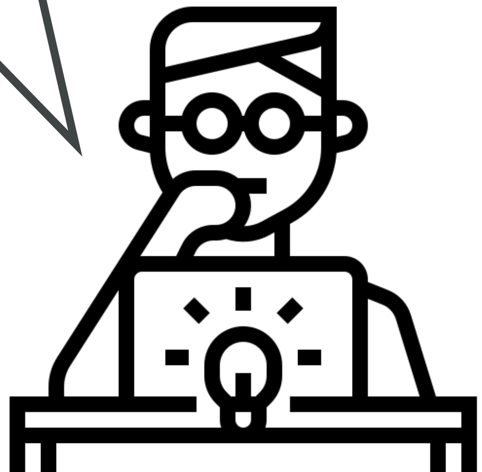
What is the probability that



is active on my target?







UQ can provide valuable information:

- for which compounds can the model **confidently** make predictions?
- about which compounds is the model **uncertain**?

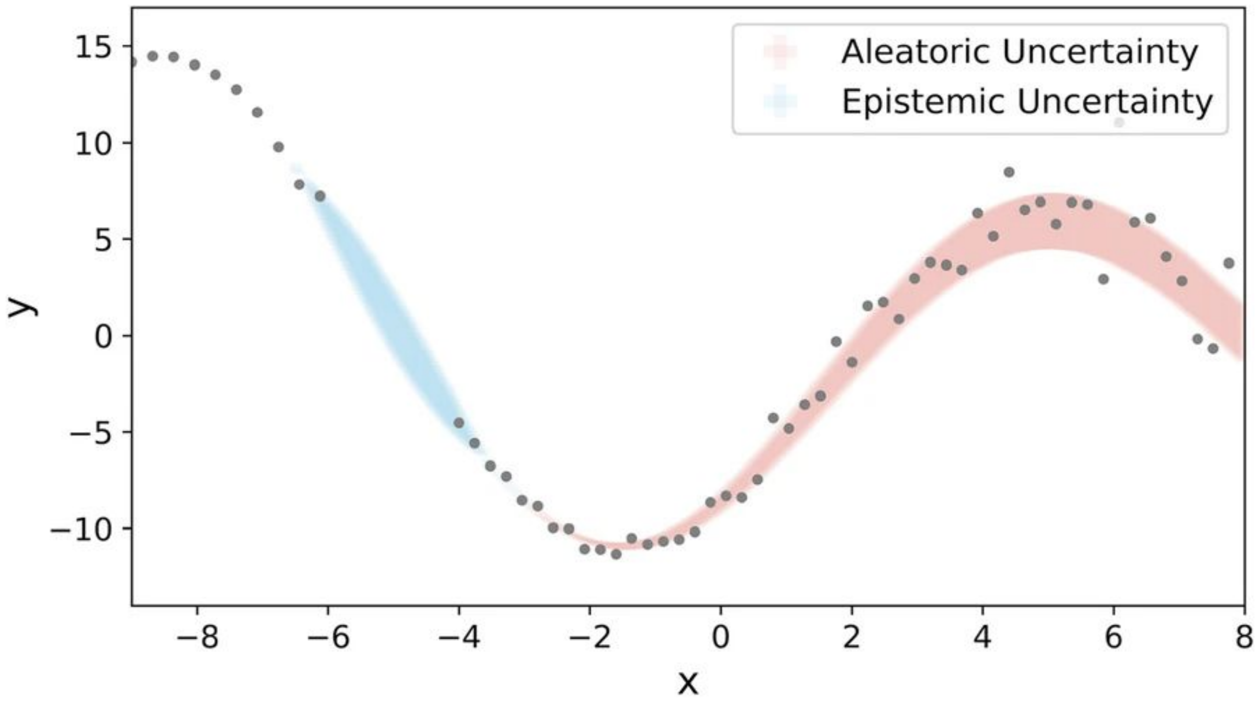
UQ can help with:

- assessment of **risks, costs and benefits**
- **prioritization** of test compounds for further analysis

Models are often poorly calibrated

Sources of Uncertainty I

Epistemic vs. Aleatoric Uncertainty



Aleatoric uncertainty:

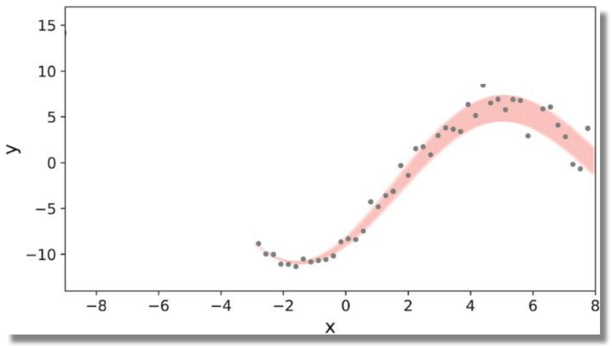
- due to **uncertainty in the data**
- stochastic/non-deterministic relationship between X and Y
- ‘irreducible’

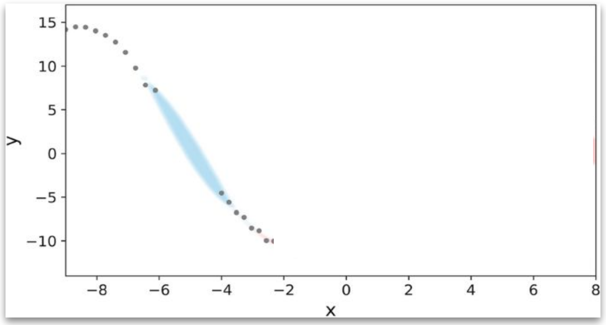
Epistemic uncertainty:

- due to **uncertainty in the model**
- everything that is not Aleatoric
- can have different sources
- ‘reducible’

Yang, Cl., Li, YP.,*J Cheminform* **15**, 13 (2023).
Hüllermeier, E., Waegeman, W. *Mach Learn* **110**, 457–506 (2021).

Sources of Uncertainty II





Uncertainty

Aleatoric
("irreducible", noise)

- **Experimental**
  (systematic, unsystematic)
- **System's property**
  (e.g. biological)

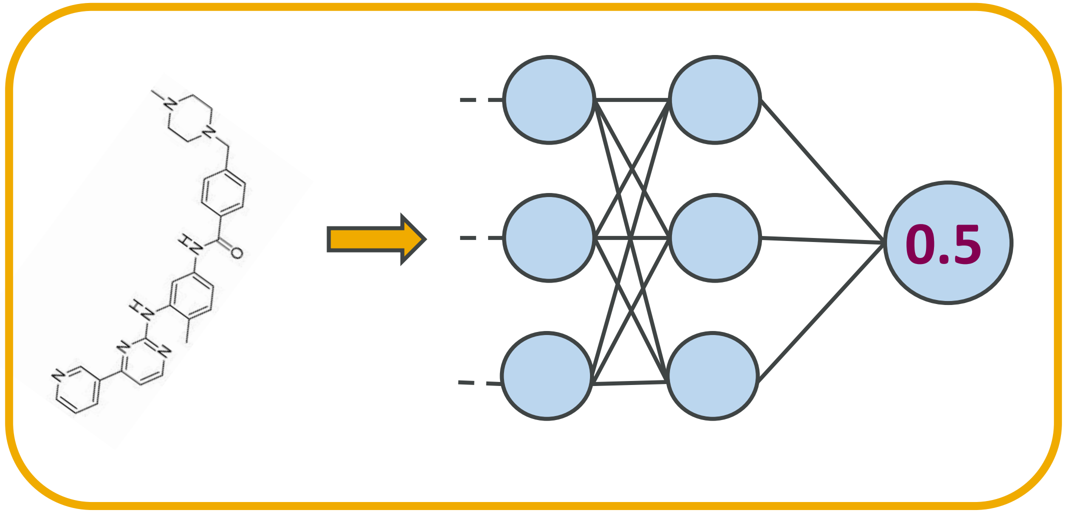
Epistemic
("reducible", lack of knowledge)

- **Model overfitting**
  Uncertainty source: Choice of Model
- **Distribution shift**
  Uncertainty source: Data

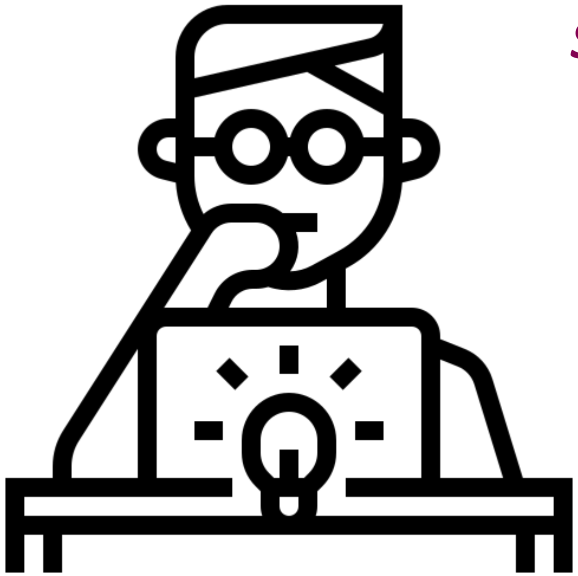
→ *leads to **increasing model variance***

Gruber, C., et al. "Sources of Uncertainty in Machine Learning--A Statisticians' View." *arXiv preprint arXiv:2305.1670*

Disentangling Uncertainty in Classification Tasks I



What does a probability of 0.5 mean?



Source of uncertainty?

Uncertainty in the data?

aleatoric?

Uncertainty in the model?

epistemic?

Disentangling Uncertainty in Classification Tasks I

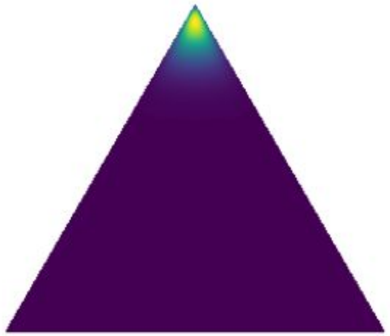
<u>Example:</u> classification task with 3 possible outcomes

Ensemble Model with M base estimators:

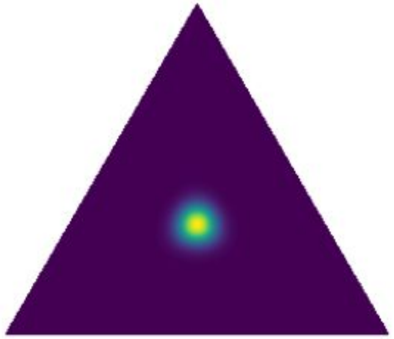
$$\{\mathrm{P}(\omega_c|\boldsymbol{x}^*, \boldsymbol{\theta}^{(i)})\}_{i=1}^{M}$$

Categorical Distributions on a Simplex

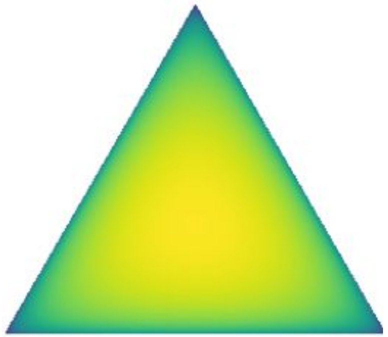
Confident Prediction



High aleatoric uncertainty
(e.g. due to experimental error)



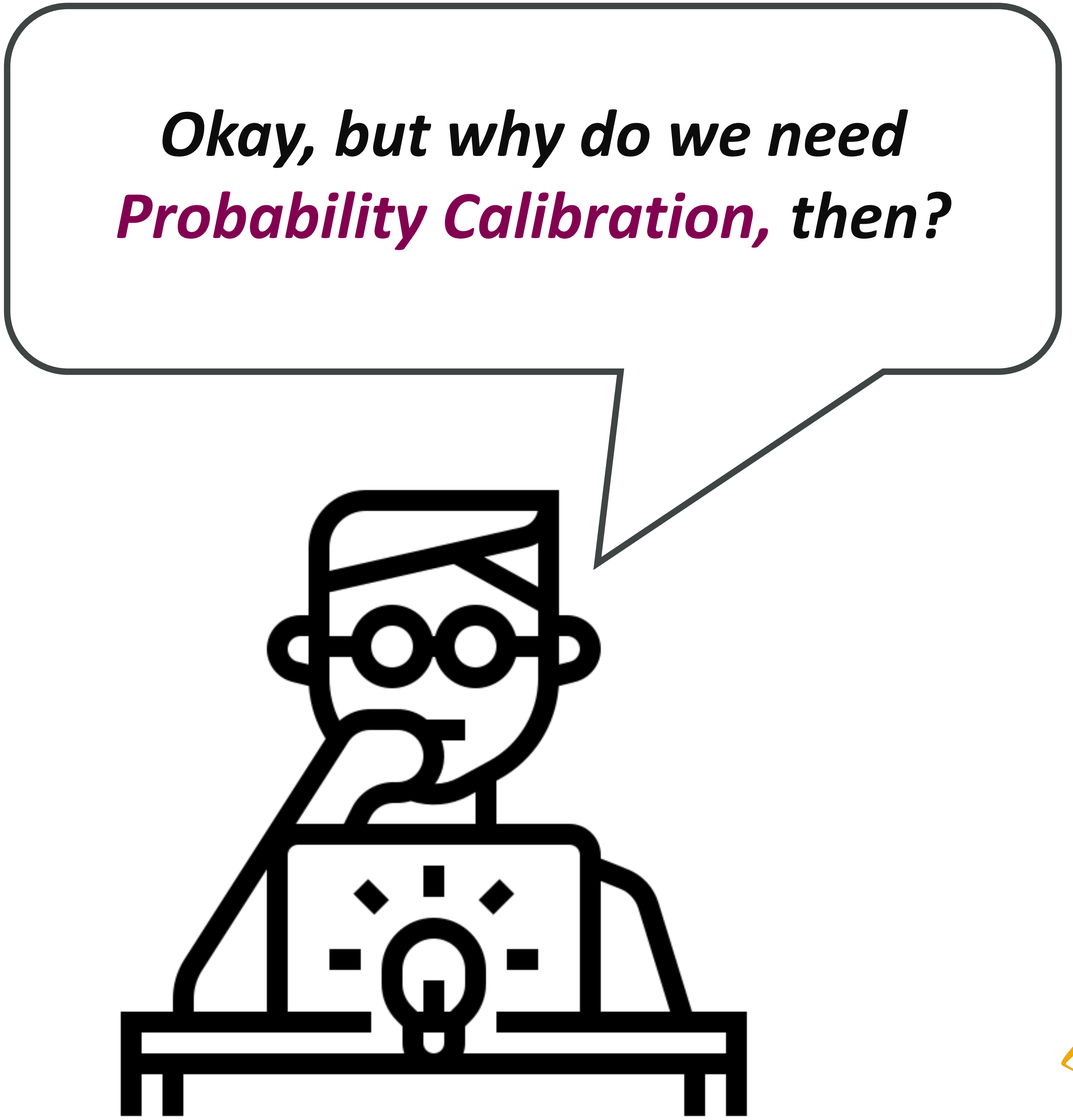
High epistemic uncertainty
(e.g. due to distribution shift)



- Predicted probability will be **the same for both cases**
- **Model variance** to determine source of uncertainty

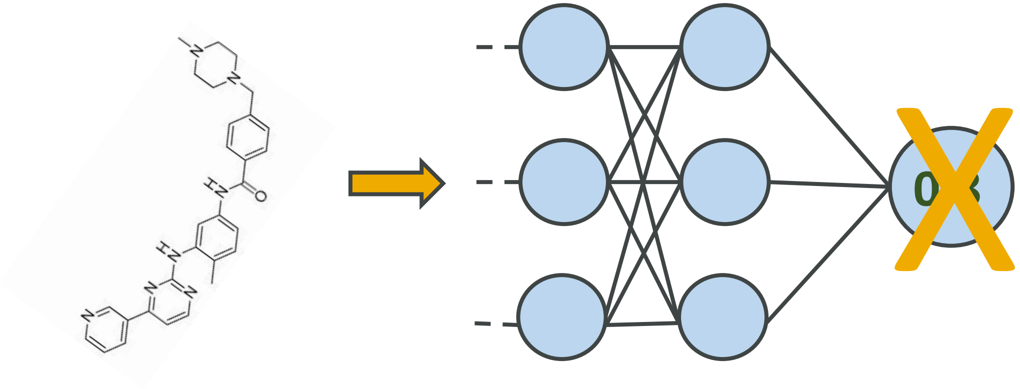
Malinin, A., and Gales M.. *Advances in neural information processing systems* 31 (2018).

Enhancing Probability Calibration in NN



Why do we need Probability Calibration (PC) ?

What if the probability is not correct?



→ **Cost assessment** cannot be performed
→ **Optimal Decision** is not feasible
→ DNN are often **poorly calibrated**

Guo, C., Pleiss, G., Sun, Y., & Weinberger, K. Q. (2017).

Why do we need Probability Calibration (PC) ?

Classification Tasks:

With **increasing** model size…

- …**classification error** will decrease
- …the model will be increasingly **miscalibrated**

During **model training**…

- …**classification error** will decrease with progressing training
- …**NLL** will first decrease and then increase with progressing training

"...the network learns better classification accuracy **at the expense of well-modeled probabilities**"

"...overfitting manifests **in probabilistic error** rather than classification error."

Guo, C., Pleiss, G., Sun, Y., & Weinberger, K. Q. (2017).

Enhancing Probability Calibration (PC) of MLPs

Chembl Data

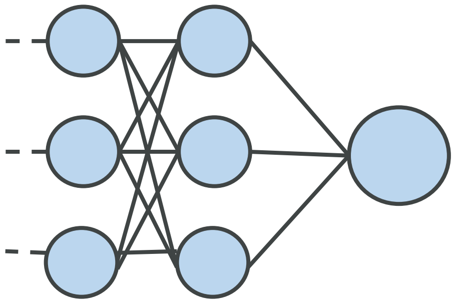
- Target: ChEMBL279

-Stratified Folds

- Subsampling Training Data

+

MLP



Enhancing Probability Calibration?

1. Model Selection:
HP-Optimization Strategies

Accuracy

ROC-AUC

✓ BCE-Loss

✓ ACE

2. Model architecture:
Probability Calibration Approaches

MC-Dopout

✓ Ensemble

✓ Platt Scaling

✓ Ensemble + Platt

Simm, J., et al. (2018). *Cell Chemical Biology.*

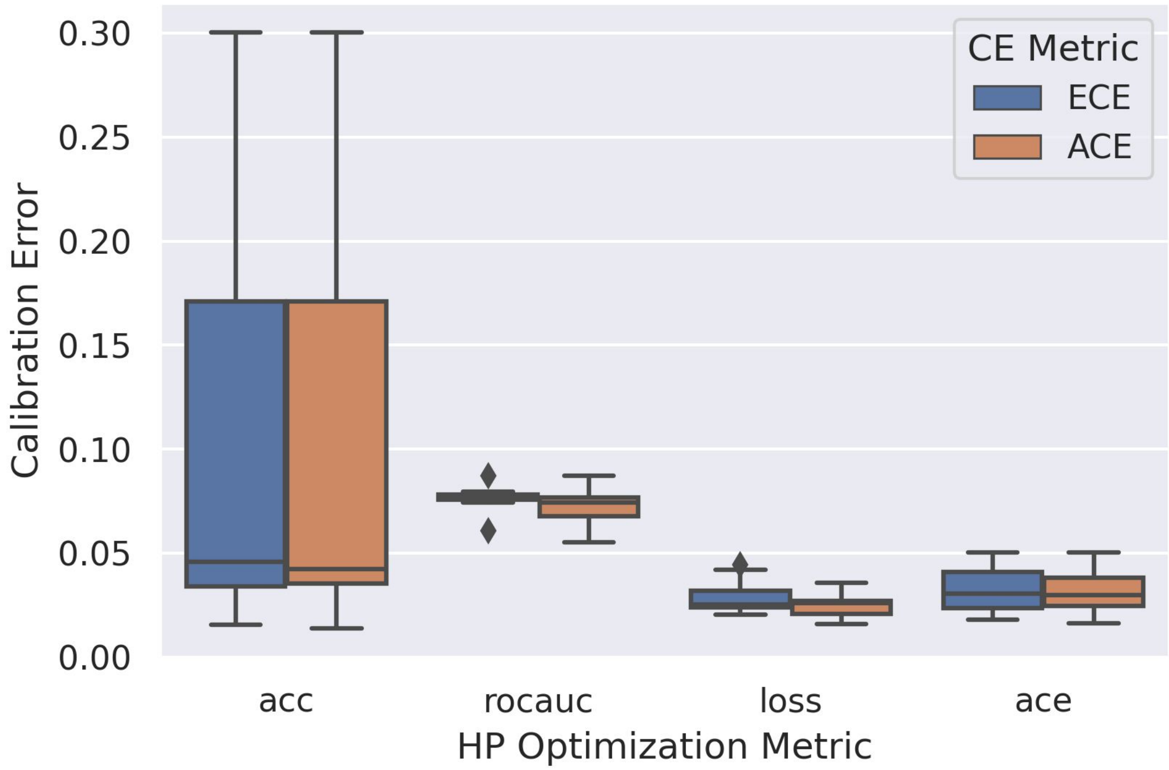
Enhancing PC: Model Selection I

Model overfitting results in **probabilistic error.**

Effects of HP- optimization strategies on PC

⟶ Compare 4 **performance metrics** used in HP- optimization

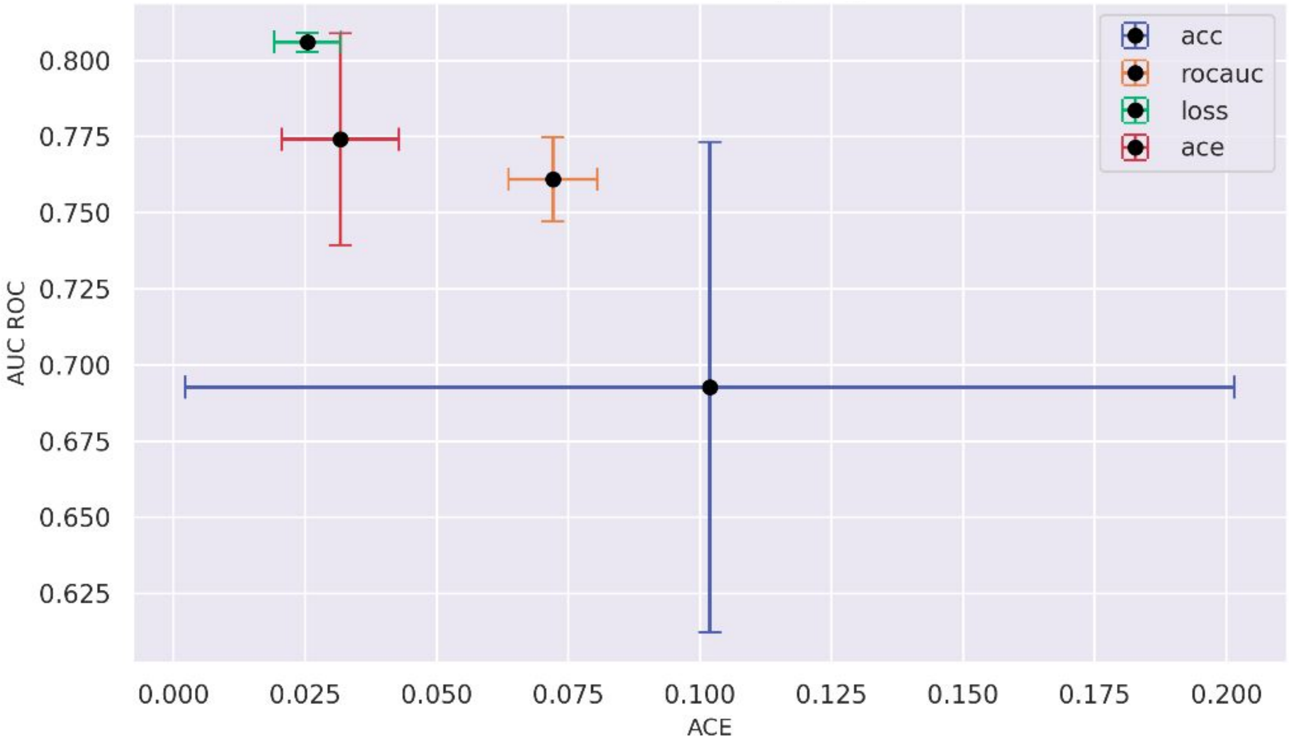
- Accuracy
- ROC-AUC
- BCE Loss
- Adaptive Calibration Error (ACE)



⟶ There are **substantial differences in CE** between the optimization strategies
⟶ Optimization **with BCE Loss or ACE** leads to the best results in PC
⟶ Both CE metics show similar results

Effects of HP- optimization Strategies on PC

How are the different Metrics performing in terms of ROC-AUC?



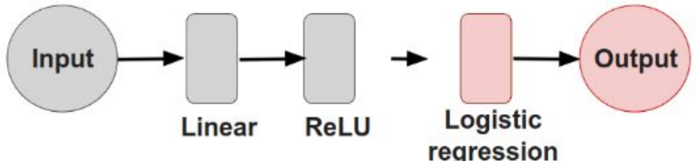
⟶ Model optimized with **Loss and Ace PC**
⟶ **Loss and Ace** perform well in terms of **ROC-AUC**
⟶ Accuracy shows the worst performance

Enhancing PC: Calibration Approaches I

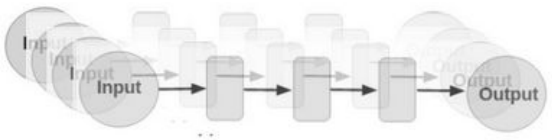
Which PC approaches help improve PC?

→ Compare 3 **calibration approaches**

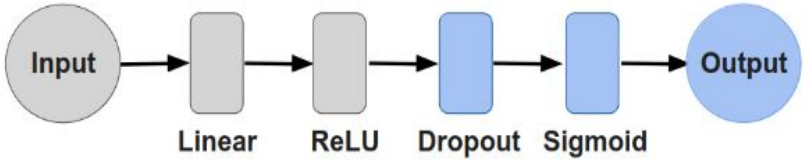
- Platt Scaling:



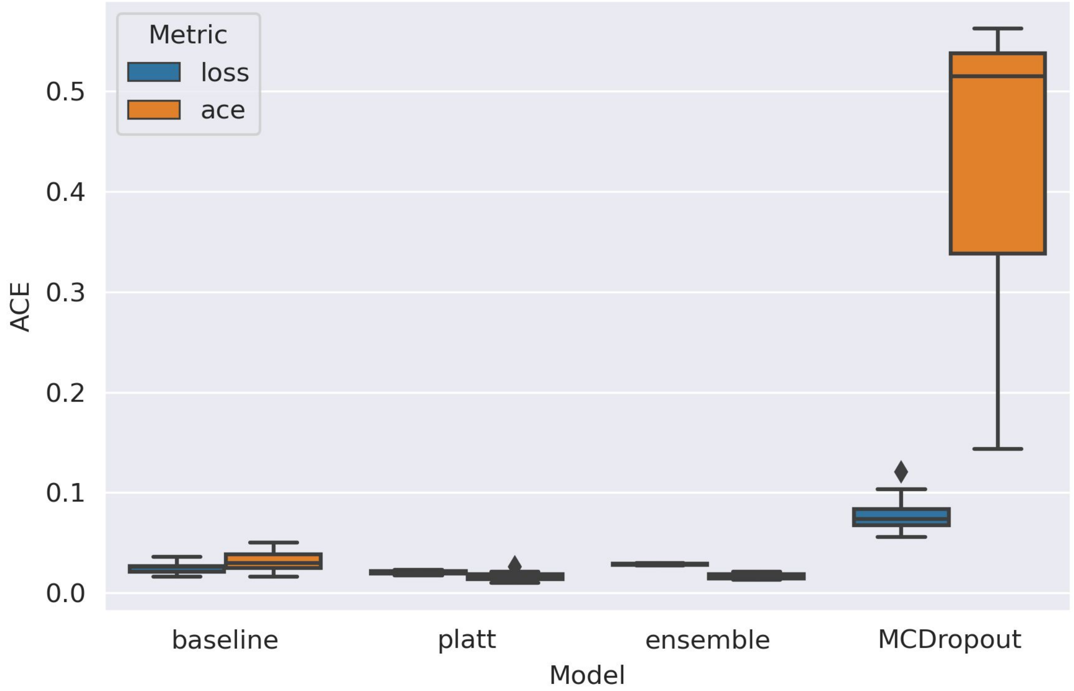
- Ensemble Model:



- MC–Dropout:



→ **HP- Optimization Metrics: Loss and ACE**



→ MC - Dropout does not work well
→ **Platt Scaling and Ensemble Modeling** look promising

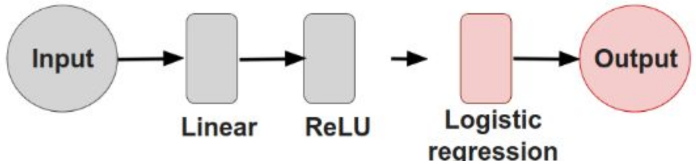
Lakshminarayanan, et al. (2017). *Advances in Neural Information Processing Systems*,.
Platt, J. C., & Platt, J. C. (1999). *Advances in Large Margin Classifiers.*

Enhancing PC: Calibration Approaches II

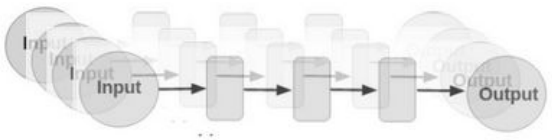
Which PC approaches help improve PC?

⟶ Compare 3 **calibration approaches**

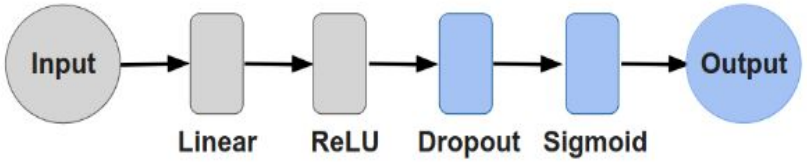
- Platt Scaling:



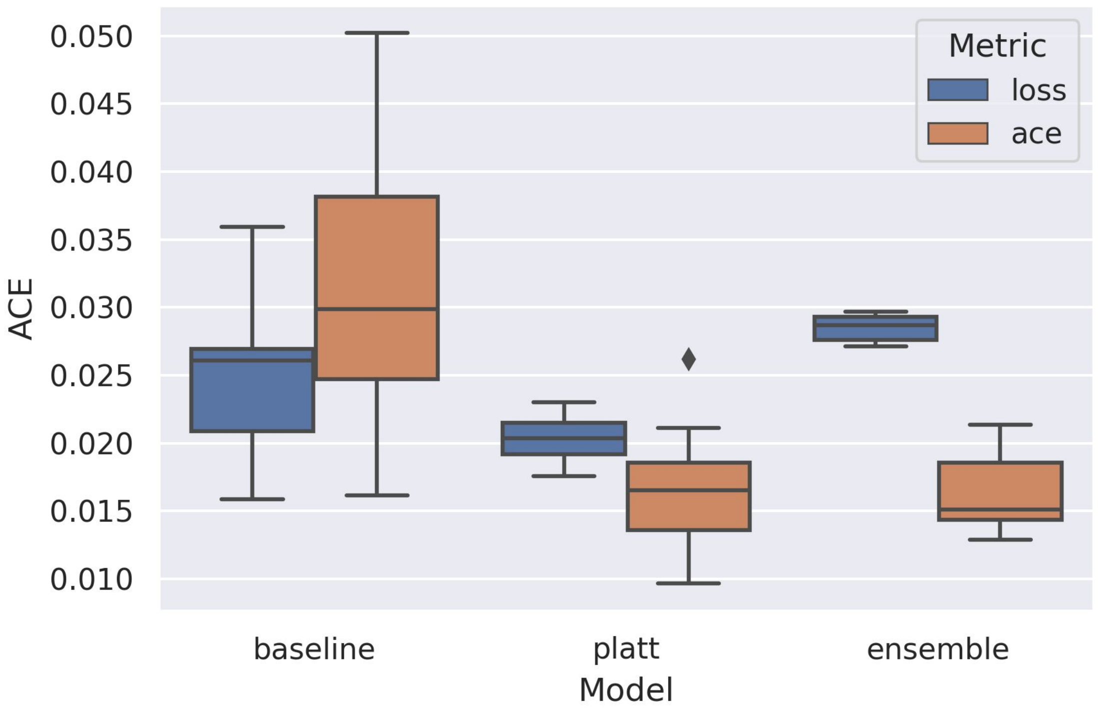
- Ensemble Model:



- MC–Dropout:



⟶ **HP- Optimization Metrics: Loss and ACE**



⟶ MC - Dropout does not work well

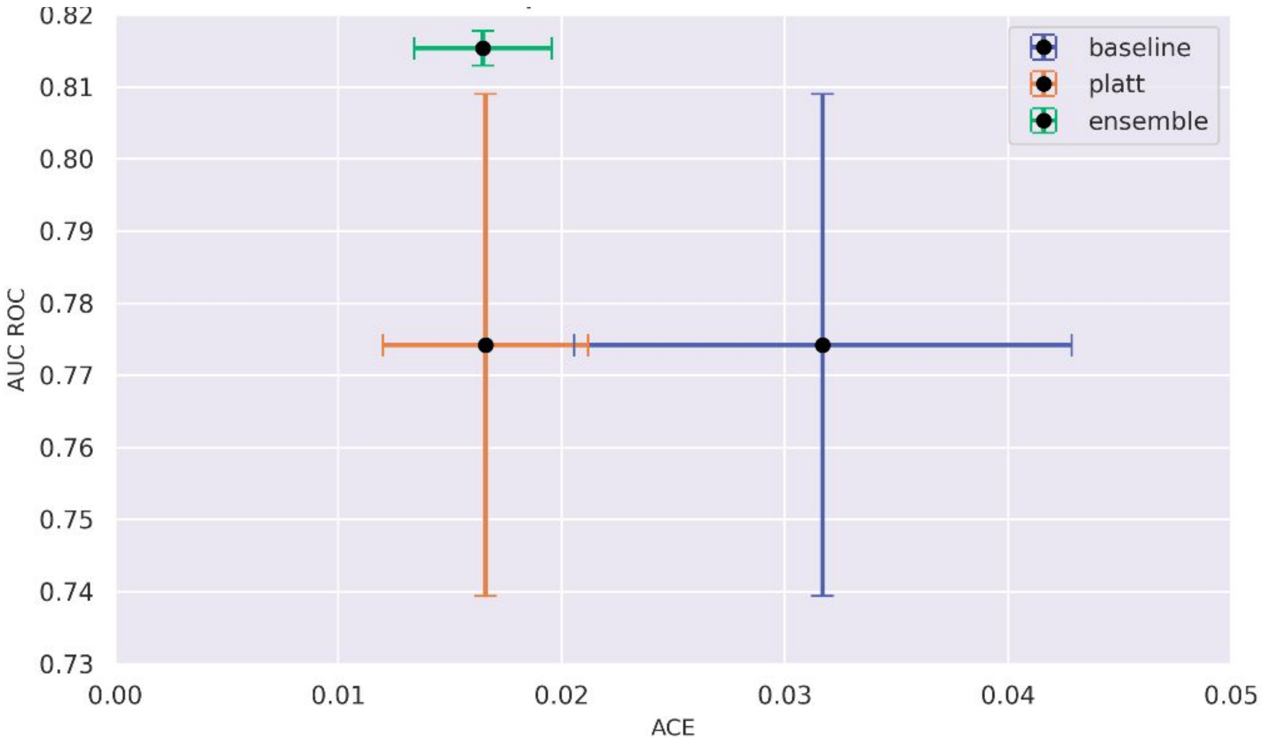
⟶ **Platt Scaling and Ensemble Modeling** improve PC

Lakshminarayanan, et al. (2017). *Advances in Neural Information Processing Systems*,.
Platt, J. C., & Platt, J. C. (1999). *Advances in Large Margin Classifiers.*

Enhancing PC: Calibration Approaches III

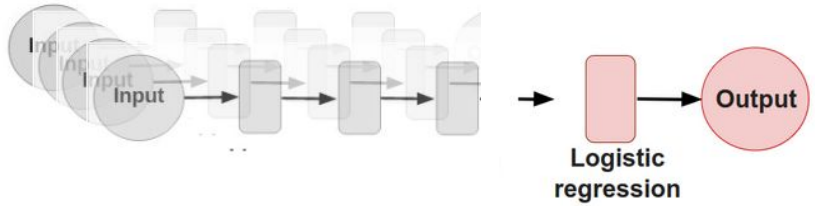
How are the Models performing in terms of ROC-AUC?

⟶ **HP- Optimization Metrics: ACE**



⟶ **Platt-Scaling and Ensemble Modeling** improve PC
⟶ Ensemble Modeling **improves ROC AUC**
⟶ Platt Scaling is **not able to improve ROC- AUC** (does not correct non/monotonous so distortions)

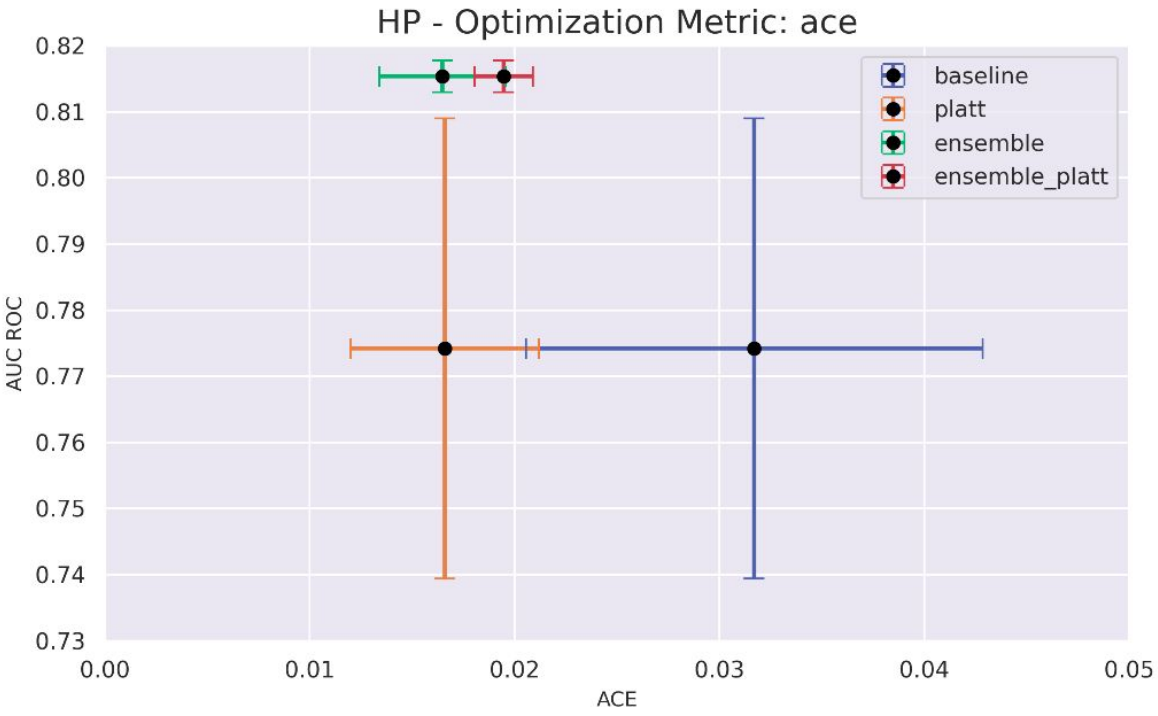
Idea: Platt- Scaling of Ensemble Model?

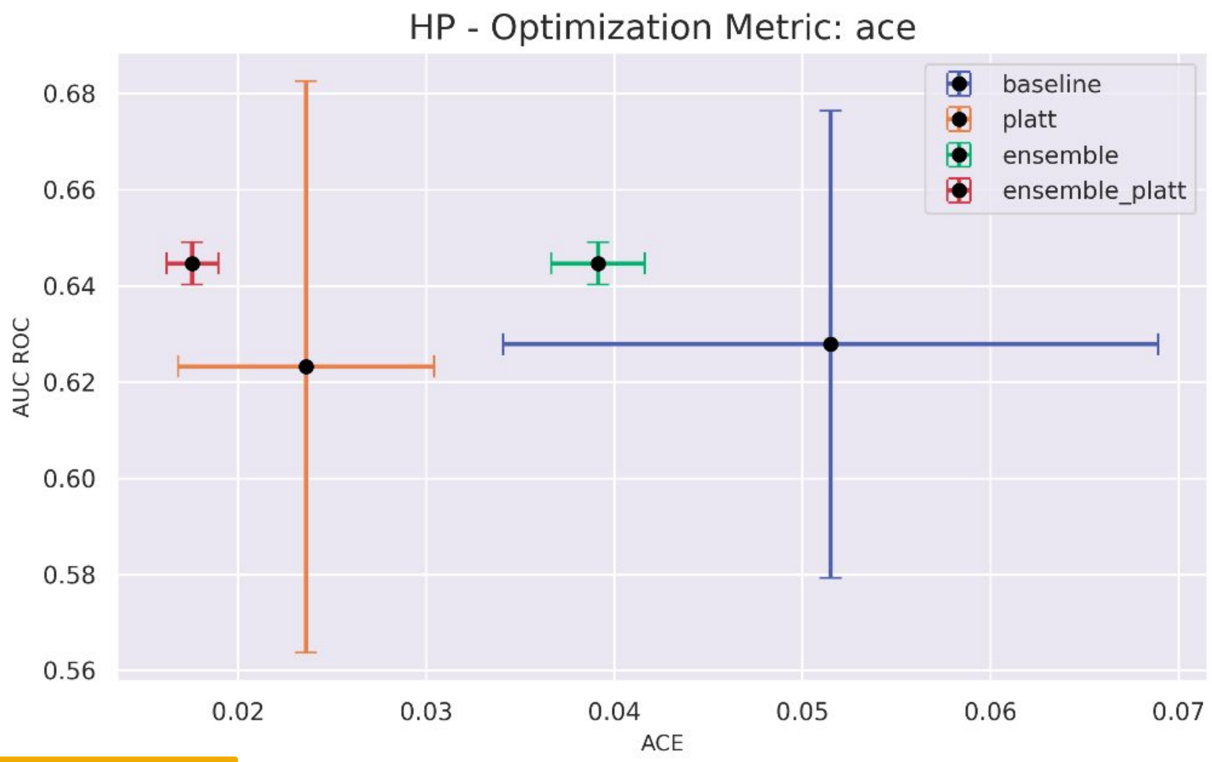


Enhancing PC: Calibration Approaches IV

Platt- Scaling of Ensemble Models:

→ **Subsampling the training data**





→ **in some cases** Platt Scaling improves PC of Ensemble Models
→ **Work in progress!**

Thank you!

References

Uncertainty estimation for Neural Networks:

- Gruber, C., et al. "Sources of Uncertainty in Machine Learning--A Statisticians' View." *arXiv preprint arXiv:2305.16703* (2023).
- Guo, C., Pleiss, G., Sun, Y., & Weinberger, K. Q. (2017). *On Calibration of Modern Neural Networks* (pp. 1321–1330). PMLR. https://proceedings.mlr.press/v70/guo17a.html
- Hüllermeier, E., Waegeman, W. Aleatoric and epistemic uncertainty in machine learning: an introduction to concepts and methods. *Mach Learn* **110**, 457–506 (2021). https://doi.org/10.1007/s10994-021-05946-3
- Malinin, A., and Gales M.. "Predictive uncertainty estimation via prior networks." *Advances in neural information processing systems* 31 (2018).
- Mervin, L. H., Johansson, S., Semenova, E., Giblin, K. A., & Engkvist, O. (2021). Uncertainty quantification in drug design. *Drug Discovery Today*, *26*(2), 474–489. https://doi.org/10.1016/J.DRUDIS.2020.11.027
- Nixon, J., Dusenberry, M., Jerfel, G., Nguyen, T., Liu, J., Zhang, L., & Tran, D. (2019). *Measuring Calibration in Deep Learning*. https://doi.org/10.48550/arxiv.1904.01685
- Yang, CI., Li, YP. Explainable uncertainty quantifications for deep learning-based molecular property prediction. *J Cheminform* **15**, 13 (2023). https://doi.org/10.1186/s13321-023-00682-3

Deep Ensembles:

- Lakshminarayanan, B., Pritzel, A., & Deepmind, C. B. (2017). Simple and Scalable Predictive Uncertainty Estimation using Deep Ensembles. *Advances in Neural Information Processing Systems*, *30*.

Platt Scaling:

- Platt, J. C., & Platt, J. C. (1999). Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods. *ADVANCES IN LARGE MARGIN CLASSIFIERS*, 61--74. http://citeseer.ist.psu.edu/viewdoc/summary?doi=10.1.1.41.1639

Clustering of data:

- Simm, J., Klambauer, G., Arany, A., Steijaert, M., Wegner, J. K., Gustin, E., Chupakhin, V., Chong, Y. T., Vialard, J., Buijnsters, P., Velter, I., Vapirev, A., Singh, S., Carpenter, A. E., Wuyts, R., Hochreiter, S., Moreau, Y., & Ceulemans, H. (2018). Repurposing High-Throughput Image Assays Enables Biological Activity Prediction for Drug Discovery. *Cell Chemical Biology*. https://doi.org/10.1016/j.chembiol.2018.01.015