Research overview

Akifumi Okuno's research concern encompasses the following four topics

A: statistical machine learning (computation),

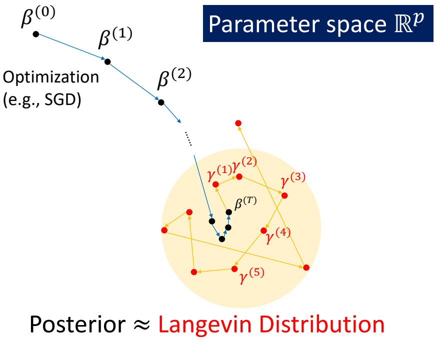
B: scientific collaborations (application),

C: mathematical statistics (theory),

D: feature learning (visualization/data integration).

A: statistical machine learning (computation)

Okuno and Yano (JCGS2023) studies a singular model extension of information criterion (called WAIC) applied to overparameterized models ($p \geq n$) including neural networks. It also proposes a Langevin-based computation, which can be implemented by existing tools (such as PyTorch) very simply.



$$\left|\mathbb{E}[\text{WAIC}] - \text{“Generalization error”}\right| \to^{\text{in prob.}} 0, \quad (p \geq n).$$

A: statistical machine learning (computation)

Okuno (arXiv:2308.02293v2) proposes a stochastic approach to train non-linear neural networks with higher-order variation regularization. It can be applied to general regression models including deep neural networks.

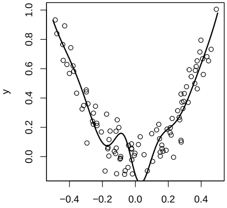


Figure: No regularization

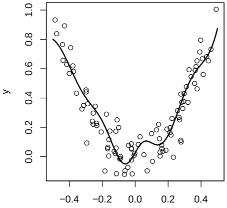


Figure: Ridge regularization

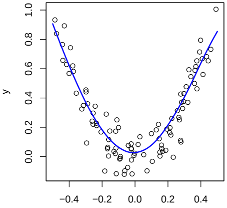
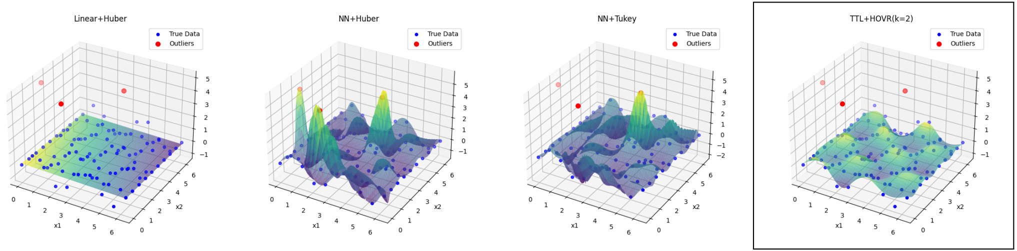


Figure: Proposed HOVR

$$\frac{1}{N}\sum_{i=1}^{N}\{y_i - f_\theta(x_i)\}^2 + \sum_{k=0}^{K}\eta_k \underbrace{\int\left\{\frac{\partial^k}{\partial x^k} f_\theta(x)\right\}^2 \mathrm{d}x}_{\text{Higher-order variation reg.}}$$

A: statistical machine learning (computation)

Okuno and Yagishita (arXiv:2308.02293v3) employs the variation regularization shown in the previous slide, to robustly train the neural network using trimmed loss. Convergence of stochastic gradient-supergradient descent (SGSD) is also proved.



(a) Linear+Huber (b) NN+Huber (c) NN+Tukey (d) **Ours**: NN+ARTL

$$\underbrace{\frac{1}{n-h}\sum_{i=1}^{n-h} r_{(i)}(\theta)^2}_{\text{Trimmed Loss}} + \underbrace{\int \left\{ \frac{\partial^k}{\partial x^k} f_\theta(x) \right\}^2 \mathrm{d}x}_{\text{Higher-order variation reg.}}$$

A: statistical machine learning (computation)

Okuno (AISM2024) proposes a stochastic approach to minimize robust divergences. It can be applied to general density models by receiving the unique benefit of stochastic outcomes, while the previous deterministic approaches can estimate only the severely restricted models (such as normal density).

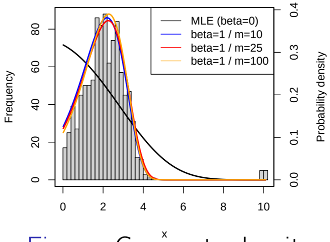


Figure: Gompertz density

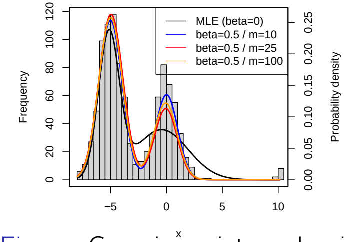
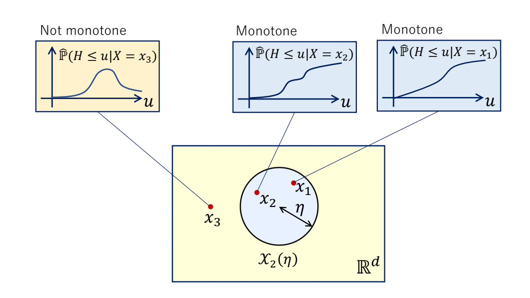


Figure: Gaussian mixture density

$$\mathbb{E}_T\left(\left\|\frac{\partial}{\partial\theta} d_\beta(\hat{Q}, P_{\theta(\tau)})\right\|_2^2\right) \to^{\text{in prob.}} 0, \quad (T \to \infty).$$

A: statistical machine learning (computation)

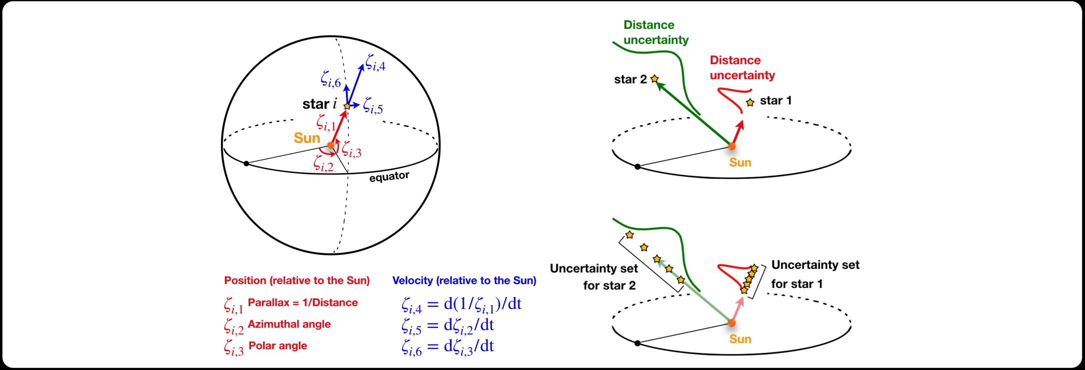
Okuno and Harada (JCGS2024) proposes a non-parallel ordinal regression model using neural networks. While existing approaches employ rather heuristic ways to preserve the monotonicity for the prediction model, it theoretically provides a sufficient condition to preserve the monotonicity.



(Sufficient condition) $$\min_{r=2,3,\ldots,R} s_{r-1} \ge \eta \cdot \rho_\infty^{[1]} \cdot \sqrt{\sum_{k=1}^{d} \left\{ \sum_{\ell=1}^{L} |w_{k,\ell}^{(2)} w_{k,\ell}^{(1)}| \right\}^2}.$$

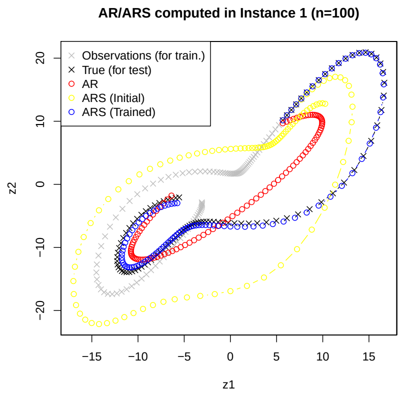
B: scientific collaborations (application)

Okuno and Hattori (arXiv:2204.08205) proposes an optimistic approach to clustering celestial bodies whose observations include large uncertainty. Subsequently, Hattori, Okuno, and Roederer (Astrophysical Journal 2023) applies it to real-world observations and obtains plausible results.



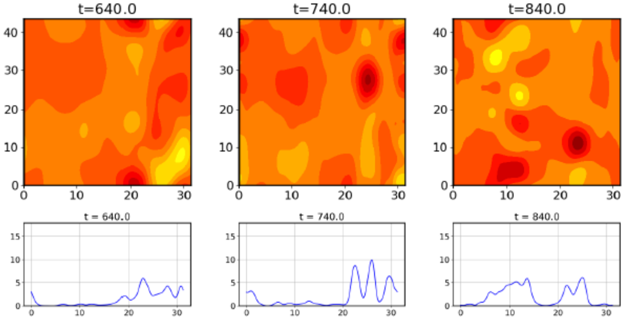
B: scientific collaborations (application)

Okuno, Morishita, and Mototake (IEEE Access 2024) proposes estimating a slack time-series for the forecasting of dynamical time-series. This very first paper of our project has been launched with the application to fusion plasmas in mind.



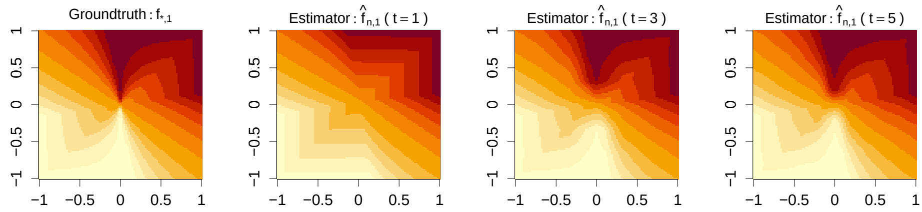
B: scientific collaborations (application)

Further extending Okuno, Kodahara, and Sasaki (Physics and Fusion Research: Rapid Communications 2024), Okuno and Sasaki (Physics of Plasmas 2025) proposes a systematic approach to decompose numerical plasma turbulence field.



C: mathematical statistics (theory)

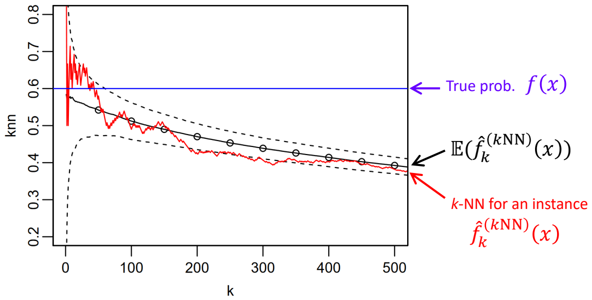
Okuno and Imaizumi (EJS2024) proves the minimax rate for the estimation of invertible functions, which have been actively developed in the field of artificial intelligence (generative models). It also provides an example estimator that achieves this minimax optimality.



$$\inf_{\hat{f}_n} \sup_{f_* \in \mathcal{F}_{\mathrm{INV}}^{\mathrm{Lip.}}} \mathrm{R}_{\mathrm{INV}}(\hat{f}_n, f_*) \asymp n^{-2/(2+d)} \quad \asymp \quad \text{minimax optimal rate}$$

C: mathematical statistics (theory)

Okuno and Shimodaira (NeurIPS2020) proves that an imaginary 0-nearest neighbor estimator corrects the higher-order bias and attains the minimax optimality. In contrast to the local polynomial regression estimator, its applicable range is broader as it can be computed with only the radial distance.



$$\mathcal{E}(\hat{g}_n) = O(n^{-2\beta/(2\beta+d)}) \quad \asymp \quad \text{minimax optimal rate (classification).}$$

C: mathematical statistics (theory)

Okuno and Yano (SPL2023) theoretically proves that the asymptotic variance of link prediction problems depends on the covariate design. Although general audiences may not be familiar, this result is surprising as the covariate design was proved not to influence the asymptotic behavior for many problems (including usual regression).

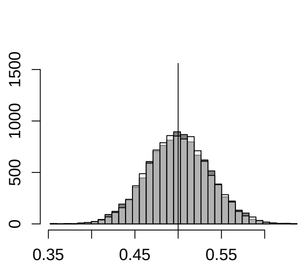


Figure: Conventional regression

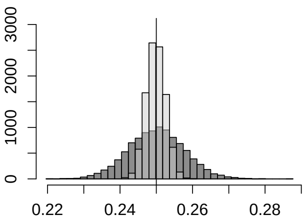
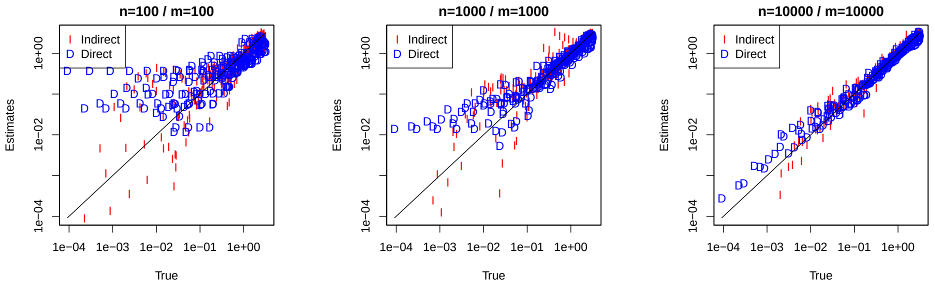


Figure: Pairwise response estimation

$$\mathbb{V}(\hat{f}_{n,h}(x, x')) = O\left(n^{-\min\{2s, 2\beta+s\}/\{s+d\}}\right).$$

C: mathematical statistics (theory)

Okuno (arXiv:2311.12380) proposes a multivariate direct kernel estimator for density ratio. It is a multivariate extension of the univariate estimator proposed by Ćwik and Mielniczuk (1989), which is defined without computing the two densities of interest.

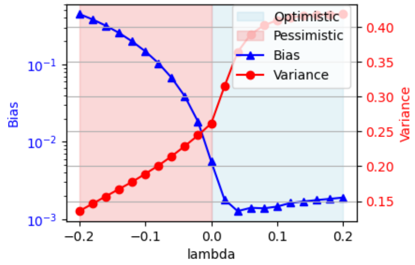


(a) $n = m = 100$ (b) $n = m = 1000$ (c) $n = m = 10000$

$$\widehat{r}_{\mathrm{Direct}}(z) := \frac{1}{nh^d} \sum_{i=1}^{n} K\left(\frac{\widehat{H}(z) - \widehat{H}(X_i)}{h}\right) \to^p r(z)$$

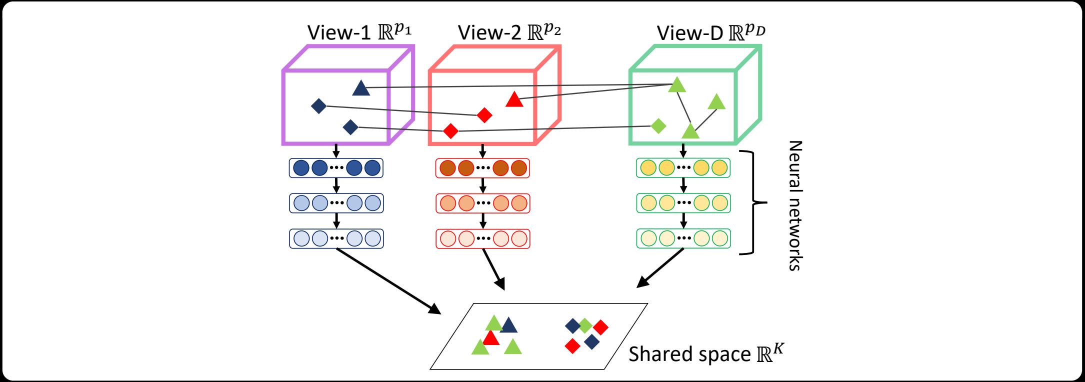
C: mathematical statistics (theory)

Okuno (arXiv:2407.10418) discusses the bias-variance trade-off that arises between robust optimization and robust statistics, highlighting that so-called "robust" methods serve distinct purposes.



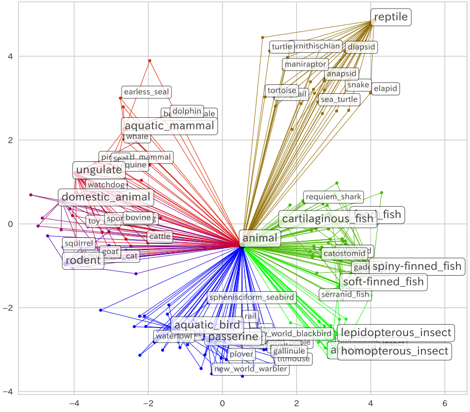
D: feature learning (visualization/data integration)

Okuno, Hada, and Shimodaira (ICML2018) proposes a probabilistic framework to integrate different types of data (texts, images,...) using their graph-structured associations (e.g., tagged information). It is further extended to hyper-relations in Okuno and Shimodaira (Neural Networks 2020).



D: feature learning (visualization/data integration)

Okuno, Kim, and Shimodaira (AISTATS2019) proposes more expressive models for neural network-based graph embedding, which reduces the dimension of data vectors having a graph-structured associations. Subsequently, Kim, Okuno, Fukui, and Shimodaira (IJCAI2019) further extends its result.



D: feature learning (visualization/data integration)

Mizutani, Okuno, Kim, and Shimodaira (arXiv:2005.00670) extends *t*-stochastic neighbor embedding (*t*-SNE) to heterogeneous data vectors (i.e., different types of data having graph-structured associations).