BIBLIOGRAPHY

[1] Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, et al. "Tensorflow: a system for large-scale machine learning." In: *Osdi*. Vol. 16. 2016. Savannah, GA, USA. 2016, pp. 265–283.

[2] Syed Mumtaz Ali and Samuel D Silvey. "A general class of coefficients of divergence of one distribution from another". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 28.1 (1966), pp. 131–142.

[3] Rohan Anil, Andrew M Dai, Orhan Firat, Melvin Johnson, Dmitry Lepikhin, Alexandre Passos, Siamak Shakeri, Emanuel Taropa, Paige Bailey, Zhifeng Chen, et al. "Palm 2 technical report". In: *arXiv preprint arXiv:2305.10403* (2023).

[4] Peter M Attia, Aditya Grover, Norman Jin, Kristen A Severson, Todor M Markov, Yang-Hung Liao, Michael H Chen, Bryan Cheong, Nicholas Perkins, Zi Yang, et al. "Closed-loop optimization of fast-charging protocols for batteries with machine learning". In: *Nature* 578.7795 (2020), pp. 397–402.

[5] Haim Avron, Vikas Sindhwani, Jiyan Yang, and Michael W Mahoney. "Quasi-Monte Carlo feature maps for shift-invariant kernels". In: *The Journal of Machine Learning Research* 17.1 (2016), pp. 4096–4133.

[6] Francis Bach. "On the equivalence between kernel quadrature rules and random feature expansions". In: *The Journal of Machine Learning Research* 18.1 (2017), pp. 714–751.

[7] Maximilian Balandat, Brian Karrer, Daniel Jiang, Samuel Daulton, Ben Letham, Andrew G Wilson, and Eytan Bakshy. "BoTorch: A framework for efficient Monte-Carlo Bayesian optimization". In: *Advances in neural information processing systems* 33 (2020), pp. 21524–21538.

[8] David J Bartholomew, Martin Knott, and Irini Moustaki. *Latent variable models and factor analysis: A unified approach*. John Wiley & Sons, 2011.

[9] Matthias Bauer, Mark van der Wilk, and Carl Edward Rasmussen. "Understanding probabilistic sparse Gaussian process approximations". In: *Advances in neural information processing systems* 29 (2016).

[10]   Thomas Bayes. "LII. An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, FRS communicated by Mr. Price, in a letter to John Canton, AMFR S". In: *Philosophical transactions of the Royal Society of London* 53 (1763), pp. 370–418.

[11]   Mohamed Ishmael Belghazi, Aristide Baratin, Sai Rajeshwar, Sherjil Ozair, Yoshua Bengio, Aaron Courville, and Devon Hjelm. "Mutual information neural estimation". In: *International conference on machine learning*. PMLR. 2018, pp. 531–540.

[12]   James O Berger. *Statistical decision theory and Bayesian analysis*. Springer Science & Business Media, 2013.

[13]   James Bergstra and Yoshua Bengio. "Random search for hyperparameter optimization". In: *The Journal of Machine Learning Research* 13.1 (2012), pp. 281–305.

[14]   James S Bergstra, Rémi Bardenet, Yoshua Bengio, and Balázs Kégl. "Algorithms for hyper-parameter optimization". In: *Advances in Neural Information Processing Systems*. 2011, pp. 2546–2554.

[15]   Steffen Bickel, Michael Brückner, and Tobias Scheffer. "Discriminative learning for differing training and test distributions". In: *Proceedings of the 24th international conference on Machine learning*. 2007, pp. 81–88.

[16]   Christopher M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Berlin, Heidelberg: Springer-Verlag, 2006. ISBN: 0387310738.

[17]   David M Blei, Alp Kucukelbir, and Jon D McAuliffe. "Variational inference: A review for statisticians". In: *Journal of the American statistical Association* 112.518 (2017), pp. 859–877.

[18]   Charles Blundell, Julien Cornebise, Koray Kavukcuoglu, and Daan Wierstra. "Weight uncertainty in neural network". In: *International conference on machine learning*. PMLR. 2015, pp. 1613–1622.

[19]   Leo Breiman. "Random forests". In: *Machine Learning* 45.1 (2001), pp. 5–32.

[20]   Eric Brochu, Vlad M Cora, and Nando De Freitas. "A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning". In: *arXiv preprint arXiv:1012.2599* (2010).

[21]   Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. "Language models are few-shot learners". In: *Advances in neural information processing systems* 33 (2020), pp. 1877–1901.

[22] Thang D Bui, Josiah Yan, and Richard E Turner. "A unifying framework for Gaussian process pseudo-point approximations using power expectation propagation". In: *The Journal of Machine Learning Research* 18.1 (2017), pp. 3649–3720.

[23] Thang Duc Bui. "Efficient Deterministic Approximate Bayesian Inference for Gaussian Process models". PhD thesis. University of Cambridge, 2018.

[24] David Burt, Carl Edward Rasmussen, and Mark Van Der Wilk. "Rates of convergence for sparse variational Gaussian process regression". In: *International Conference on Machine Learning*. PMLR. 2019, pp. 862–871.

[25] David R Burt, Carl Edward Rasmussen, and Mark van der Wilk. "Variational orthogonal features". In: *arXiv preprint arXiv:2006.13170* (2020).

[26] Richard H Byrd, Peihuang Lu, Jorge Nocedal, and Ciyou Zhu. "A limited memory algorithm for bound constrained optimization". In: *SIAM Journal on scientific computing* 16.5 (1995), pp. 1190–1208.

[27] Russel E Caflisch. "Monte carlo and quasi-monte carlo methods". In: *Acta numerica* 7 (1998), pp. 1–49.

[28] Roberto Calandra, Jan Peters, Carl Edward Rasmussen, and Marc Peter Deisenroth. "Manifold Gaussian processes for regression". In: *2016 International joint conference on neural networks (IJCNN)*. IEEE. 2016, pp. 3338–3345.

[29] Daniele Calandriello, Luigi Carratino, Alessandro Lazaric, Michal Valko, and Lorenzo Rosasco. "Gaussian process optimization with adaptive sketching: Scalable and no regret". In: *Conference on Learning Theory*. PMLR. 2019, pp. 533–557.

[30] Peter E Castro, W H_ Lawton, and EA Sylvestre. "Principal modes of variation for processes with continuous sample curves". In: *Technometrics* 28.4 (1986), pp. 329–337.

[31] Liqun Chen, Shuyang Dai, Yunchen Pu, Erjin Zhou, Chunyuan Li, Qinliang Su, Changyou Chen, and Lawrence Carin. "Symmetric variational autoencoder and connections to adversarial learning". In: *International Conference on Artificial Intelligence and Statistics*. PMLR. 2018, pp. 661–669.

[32] Peng Chen, Brian M Merrick, and Thomas J Brazil. "Bayesian optimization for broadband high-efficiency power amplifier designs". In: *IEEE Transactions on Microwave Theory and Techniques* 63.12 (2015), pp. 4263–4272.

[33] Tianqi Chen, Emily Fox, and Carlos Guestrin. "Stochastic gradient hamiltonian monte carlo". In: *International conference on machine learning*. PMLR. 2014, pp. 1683–1691.

[34] Tianqi Chen and Carlos Guestrin. "Xgboost: A scalable tree boosting system". In: *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. 2016, pp. 785–794.

[35] Xi Chen, Yan Duan, Rein Houthooft, John Schulman, Ilya Sutskever, and Pieter Abbeel. "Infogan: Interpretable representation learning by information maximizing generative adversarial nets". In: *Advances in neural information processing systems* 29 (2016).

[36] Ching-An Cheng and Byron Boots. "Variational inference for Gaussian process models with linear complexity". In: *Advances in Neural Information Processing Systems* 30 (2017).

[37] Kuang Fu Cheng, Chih-Kang Chu, et al. "Semiparametric density estimation under a two-sample density ratio model". In: *Bernoulli* 10.4 (2004), pp. 583–604.

[38] Kristy Choi, Aditya Grover, Trisha Singh, Rui Shu, and Stefano Ermon. "Fair generative modeling via weak supervision". In: *International Conference on Machine Learning*. PMLR. 2020, pp. 1887–1898.

[39] Francois Chollet. *Deep learning with Python*. Simon and Schuster, 2021.

[40] Krzysztof Choromanski, Mark Rowland, Tamás Sarlós, Vikas Sindhwani, Richard Turner, and Adrian Weller. "The geometry of random features". In: *International Conference on Artificial Intelligence and Statistics*. PMLR. 2018, pp. 1–9.

[41] Krzysztof Choromanski, Mark Rowland, and Adrian Weller. "The unreasonable effectiveness of structured random orthogonal embeddings". In: *Proceedings of the 31st International Conference on Neural Information Processing Systems*. 2017, pp. 218–227.

[42] Patryk Chrabaszcz, Ilya Loshchilov, and Frank Hutter. "A downsampled variant of imagenet as an alternative to the cifar datasets". In: *arXiv preprint arXiv:1707.08819* (2017).

[43] Djork-Arné Clevert, Thomas Unterthiner, and Sepp Hochreiter. "Fast and accurate deep network learning by exponential linear units (elus)". In: *arXiv preprint arXiv:1511.07289* (2015).

[44] Dorin Comaniciu and Peter Meer. "Mean shift: A robust approach toward feature space analysis". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24.5 (2002), pp. 603–619.

[45] Lehel Csató and Manfred Opper. "Sparse on-line Gaussian processes". In: *Neural computation* 14.3 (2002), pp. 641–668.

[46] Imre Csiszár. "On information-type measure of difference of probability distributions and indirect observations". In: *Studia Sci. Math. Hungar.* 2 (1967), pp. 299–318.

[47] Imre Csiszár, Paul C Shields, et al. "Information theory and statistics: A tutorial". In: *Foundations and Trends® in Communications and Information Theory* 1.4 (2004), pp. 417–528.

[48] Andreas Damianou and Neil D Lawrence. "Deep gaussian processes". In: *Artificial intelligence and statistics*. PMLR. 2013, pp. 207–215.

[49] Tri Dao, Christopher De Sa, and Christopher Ré. "Gaussian quadrature for kernel features". In: *Advances in neural information processing systems* 30 (2017), p. 6109.

[50] Philip J Davis and Philip Rabinowitz. *Methods of numerical integration*. Courier Corporation, 2007.

[51] Peter Dayan, Geoffrey E Hinton, Radford M Neal, and Richard S Zemel. "The helmholtz machine". In: *Neural computation* 7.5 (1995), pp. 889–904.

[52] George De Ath, Tinkle Chugh, and Alma AM Rahat. "MBORE: multi-objective Bayesian optimisation by density-ratio estimation". In: *Proceedings of the Genetic and Evolutionary Computation Conference*. 2022, pp. 776–785.

[53] Morris H DeGroot. *Optimal statistical decisions*. John Wiley & Sons, 2005.

[54] Marc Deisenroth and Carl E Rasmussen. "PILCO: A model-based and data-efficient approach to policy search". In: *Proceedings of the 28th International Conference on machine learning (ICML-11)*. 2011, pp. 465–472.

[55] Marc Peter Deisenroth, A Aldo Faisal, and Cheng Soon Ong. *Mathematics for machine learning*. Cambridge University Press, 2020.

[56] Amir Dezfouli and Edwin V Bonilla. "Scalable inference for Gaussian process models with black-box likelihoods". In: *Advances in Neural Information Processing Systems* 28 (2015).

[57] Josef Dick, Frances Y Kuo, and Ian H Sloan. "High-dimensional integration: the quasi-Monte Carlo way". In: *Acta Numerica* 22 (2013), pp. 133–288.

[58] Jeff Donahue, Philipp Krähenbühl, and Trevor Darrell. "Adversarial feature learning". In: *arXiv preprint arXiv:1605.09782* (2016).

[59] Xuanyi Dong and Yi Yang. "Nas-bench-102: Extending the scope of reproducible neural architecture search". In: *arXiv preprint arXiv:2001.00326* (2020).

[60] Dheeru Dua and Casey Graff. *UCI Machine Learning Repository*. 2017. URL: http://archive.ics.uci.edu/ml.

[61] Vincent Dumoulin, Ishmael Belghazi, Ben Poole, Olivier Mastropietro, Alex Lamb, Martin Arjovsky, and Aaron Courville. "Adversarially learned inference". In: *arXiv preprint arXiv:1606.00704* (2016).

[62] Joseph Duris, Dylan Kennedy, Adi Hanuka, Jane Shtalenkova, Auralee Edelen, P Baxevanis, Adam Egger, T Cope, M McIntire, S Ermon, et al. "Bayesian optimization of a free-electron laser". In: *Physical review letters* 124.12 (2020), p. 124801.

[63] Conor Durkan, Iain Murray, and George Papamakarios. "On contrastive learning for likelihood-free inference". In: *International conference on machine learning*. PMLR. 2020, pp. 2771–2781.

[64] Vincent Dutordoir, Nicolas Durrande, and James Hensman. "Sparse Gaussian processes with spherical harmonic features". In: *International Conference on Machine Learning*. PMLR. 2020, pp. 2793–2802.

[65] Vincent Dutordoir, James Hensman, Mark van der Wilk, Carl Henrik Ek, Zoubin Ghahramani, and Nicolas Durrande. "Deep neural networks as point estimates for deep Gaussian processes". In: *Advances in Neural Information Processing Systems* 34 (2021).

[66] Costas Efthimiou and Christopher Frye. *Spherical harmonics in p dimensions*. World Scientific, 2014.

[67] Stefan Falkner, Aaron Klein, and Frank Hutter. "BOHB: Robust and Efficient Hyperparameter Optimization at Scale". In: *International Conference on Machine Learning*. 2018, pp. 1437–1446.

[68] Louis Napoleon George Filon. "III.–On a quadrature formula for trigonometric integrals". In: *Proceedings of the Royal Society of Edinburgh* 49 (1928), pp. 38–47.

[69] Alexander IJ Forrester and Andy J Keane. "Recent advances in surrogate-based optimization". In: *Progress in aerospace sciences* 45.1-3 (2009), pp. 50–79.

[70] Peter I Frazier. "A tutorial on Bayesian optimization". In: *arXiv preprint arXiv:1807.02811* (2018).

[71] Brendan J Frey and Geoffrey E Hinton. "Variational learning in nonlinear Gaussian belief networks". In: *Neural Computation* 11.1 (1999), pp. 193–213.

[72] Keinosuke Fukunaga. *Introduction to statistical pattern recognition*. Elsevier, 2013.

[73]   Yarin Gal and Zoubin Ghahramani. "Dropout as a bayesian approximation: Representing model uncertainty in deep learning". In: *international conference on machine learning*. PMLR. 2016, pp. 1050–1059.

[74]   Roman Garnett. *Bayesian Optimization*. to appear. Cambridge University Press, 2023.

[75]   Roman Garnett, Michael A Osborne, and Stephen J Roberts. "Bayesian optimization for sensor set selection". In: *Proceedings of the 9th ACM/IEEE international conference on information processing in sensor networks*. 2010, pp. 209–219.

[76]   Eduardo C Garrido-Merchán and Daniel Hernández-Lobato. "Dealing with categorical and integer-valued variables in Bayesian optimization with Gaussian processes". In: *Neurocomputing* 380 (2020), pp. 20–35.

[77]   Carl Friedrich Gauss. *Methodvs nova integralivm valores per approximationem inveniendi*. apvd Henricvm Dieterich, 1815.

[78]   Walter Gautschi. "Construction of Gauss-Christoffel quadrature formulas". In: *Mathematics of Computation* 22.102 (1968), pp. 251–270.

[79]   Walter Gautschi. "A survey of Gauss-Christoffel quadrature formulae". In: *EB Christoffel*. Springer, 1981, pp. 72–147.

[80]   Andrew Gelman, John B Carlin, Hal S Stern, David B Dunson, Aki Vehtari, and Donald B Rubin. *Bayesian data analysis*. CRC press, 2013.

[81]   Samuel Gershman and Noah Goodman. "Amortized inference in probabilistic reasoning". In: *Proceedings of the annual meeting of the cognitive science society*. Vol. 36. 2014.

[82]   Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. "Rich feature hierarchies for accurate object detection and semantic segmentation". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2014, pp. 580–587.

[83]   Tilmann Gneiting and Adrian E Raftery. "Strictly proper scoring rules, prediction, and estimation". In: *Journal of the American Statistical Association* 102.477 (2007), pp. 359–378.

[84]   Ian J Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. "Generative Adversarial Networks". In: *arXiv preprint arXiv:1406.2661* (2014).

[85]   Alex Graves, Abdel-rahman Mohamed, and Geoffrey Hinton. "Speech recognition with deep recurrent neural networks". In: *2013 IEEE international conference on acoustics, speech and signal processing*. Ieee. 2013, pp. 6645–6649.

[86]   Arthur Gretton, Alex Smola, Jiayuan Huang, Marcel Schmitt-full, Karsten Borgwardt, and Bernhard Schölkopf. "Covariate shift by kernel mean matching". In: *Dataset Shift in Machine Learning* 3.4 (2009), p. 5.

[87]   Ryan-Rhys Griffiths, Alexander A Aldrick, Miguel Garcia-Ortegon, Vidhi Lalchand, et al. "Achieving robustness to aleatoric uncertainty with heteroscedastic Bayesian optimisation". In: *Machine Learning: Science and Technology* 3.1 (2021), p. 015004.

[88]   Aditya Grover and Stefano Ermon. "Boosted generative models". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 32. 1. 2018.

[89]   Aditya Grover, Jiaming Song, Ashish Kapoor, Kenneth Tran, Alekh Agarwal, Eric J Horvitz, and Stefano Ermon. "Bias correction of learned generative models using likelihood-free importance weighting". In: *Advances in neural information processing systems* 32 (2019).

[90]   Michael U Gutmann and Aapo Hyvärinen. "Noise-Contrastive Estimation of Unnormalized Statistical Models, with Applications to Natural Image Statistics." In: *Journal of machine learning research* 13.2 (2012).

[91]   Heikki Haario, Eero Saksman, and Johanna Tamminen. "Adaptive proposal distribution for random walk Metropolis algorithm". In: *Computational statistics* 14 (1999), pp. 375–395.

[92]   Philipp Hennig, Michael A Osborne, and Hans P Kersting. *Probabilistic Numerics*. Cambridge University Press, 2022.

[93]   Philipp Hennig and Christian J Schuler. "Entropy Search for Information-Efficient Global Optimization." In: *Journal of Machine Learning Research* 13.6 (2012).

[94]   James Hensman, Nicolas Durrande, and Arno Solin. "Variational Fourier Features for Gaussian Processes". In: *Journal of Machine Learning Research* 18.151 (2018), pp. 1–52. URL: http://jmlr.org/papers/v18/16-579.html.

[95]   James Hensman, Nicolo Fusi, and Neil D Lawrence. "Gaussian processes for big data". In: *arXiv preprint arXiv:1309.6835* (2013).

[96]   James Hensman, Alexander Matthews, and Zoubin Ghahramani. "Scalable variational Gaussian process classification". In: *Artificial Intelligence and Statistics*. PMLR. 2015, pp. 351–360.

[97]   James Hensman, Alexander G Matthews, Maurizio Filippone, and Zoubin Ghahramani. "MCMC for variationally sparse Gaussian processes". In: *Advances in Neural Information Processing Systems* 28 (2015).

[98] Daniel Hernández-Lobato, Jose Hernandez-Lobato, Amar Shah, and Ryan Adams. "Predictive entropy search for multi-objective Bayesian optimization". In: *International Conference on Machine Learning.* 2016, pp. 1492–1501.

[99] José Miguel Hernández-Lobato, Matthew W Hoffman, and Zoubin Ghahramani. "Predictive entropy search for efficient global optimization of black-box functions". In: *Advances in neural information processing systems* 27 (2014).

[100] Geoffrey Hinton, Li Deng, Dong Yu, George E Dahl, Abdelrahman Mohamed, Navdeep Jaitly, Andrew Senior, Vincent Vanhoucke, Patrick Nguyen, Tara N Sainath, et al. "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups". In: *IEEE Signal processing magazine* 29.6 (2012), pp. 82–97.

[101] Jonathan Ho, Ajay Jain, and Pieter Abbeel. "Denoising diffusion probabilistic models". In: *Advances in neural information processing systems* 33 (2020), pp. 6840–6851.

[102] Matthew D Hoffman, David M Blei, Chong Wang, and John Paisley. "Stochastic variational inference". In: *Journal of Machine Learning Research* (2013).

[103] Jordan Hoffmann, Sebastian Borgeaud, Arthur Mensch, Elena Buchatskaya, Trevor Cai, Eliza Rutherford, Diego de Las Casas, Lisa Anne Hendricks, Johannes Welbl, Aidan Clark, et al. "Training compute-optimal large language models". In: *arXiv preprint arXiv:2203.15556* (2022).

[104] Kurt Hornik, Maxwell Stinchcombe, Halbert White, et al. "Multilayer feedforward networks are universal approximators." In: *Neural Networks* 2.5 (1989), pp. 359–366.

[105] Neil Houlsby, Ferenc Huszár, Zoubin Ghahramani, and Máté Lengyel. "Bayesian active learning for classification and preference learning". In: *arXiv preprint arXiv:1112.5745* (2011).

[106] Zhiting Hu, Zichao Yang, Ruslan Salakhutdinov, and Eric P Xing. "On unifying deep generative models". In: *arXiv preprint arXiv:1706.00550* (2017).

[107] Ferenc Huszár. "Variational inference using implicit distributions". In: *arXiv preprint arXiv:1702.08235* (2017).

[108] Frank Hutter, Holger H Hoos, and Kevin Leyton-Brown. "Sequential model-based optimization for general algorithm configuration". In: *International Conference on Learning and Intelligent Optimization.* Springer. 2011, pp. 507–523.

[109] Carl Gustav Jakob Jacobi. "Über Gauss neue Methode, die Werthe der Integrale näherungsweise zu finden." In: (1826).

[110]   Achin Jain and Manfred Morari. "Computing the racing line using Bayesian optimization". In: *2020 59th IEEE Conference on Decision and Control (CDC)*. IEEE. 2020, pp. 6192–6197.

[111]   Eric Jang, Shixiang Gu, and Ben Poole. "Categorical reparameterization with gumbel-softmax". In: *arXiv preprint arXiv:1611.01144* (2016).

[112]   Edwin T Jaynes. "Prior probabilities". In: *IEEE Transactions on systems science and cybernetics* 4.3 (1968), pp. 227–241.

[113]   Rodolphe Jenatton, Cedric Archambeau, Javier González, and Matthias Seeger. "Bayesian optimization with tree-structured dependencies". In: *International Conference on Machine Learning*. PMLR. 2017, pp. 1655–1664.

[114]   Donald R Jones, Matthias Schonlau, and William J Welch. "Efficient global optimization of expensive black-box functions". In: *Journal of Global Optimization* 13.4 (1998), pp. 455–492.

[115]   Michael I Jordan, Zoubin Ghahramani, Tommi S Jaakkola, and Lawrence K Saul. "An introduction to variational methods for graphical models". In: *Learning in graphical models* (1998), pp. 105–161.

[116]   Michael I Jordan, Zoubin Ghahramani, Tommi S Jaakkola, and Lawrence K Saul. "An introduction to variational methods for graphical models". In: *Machine learning* 37 (1999), pp. 183–233.

[117]   Andre G Journel and Charles J Huijbregts. "Mining geostatistics". In: (1976).

[118]   John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, et al. "Highly accurate protein structure prediction with AlphaFold". In: *Nature* 596.7873 (2021), pp. 583–589.

[119]   Takafumi Kanamori, Shohei Hido, and Masashi Sugiyama. "A least-squares approach to direct importance estimation". In: *The Journal of Machine Learning Research* 10 (2009), pp. 1391–1445.

[120]   Takafumi Kanamori, Taiji Suzuki, and Masashi Sugiyama. "Theoretical analysis of density ratio estimation". In: *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences* 93.4 (2010), pp. 787–798.

[121]   Kirthevasan Kandasamy, Gautam Dasarathy, Jeff Schneider, and Barnabás Póczos. "Multi-fidelity Bayesian optimisation with continuous approximations". In: *arXiv preprint arXiv:1703.06240* (2017).

[122] Taeksoo Kim, Moonsu Cha, Hyunsoo Kim, Jung Kwon Lee, and Jiwon Kim. "Learning to discover cross-domain relations with generative adversarial networks". In: *International conference on machine learning*. PMLR. 2017, pp. 1857–1865.

[123] Diederik P Kingma and Jimmy Ba. "Adam: A method for stochastic optimization". In: *arXiv preprint arXiv:1412.6980* (2014).

[124] Diederik P Kingma and Max Welling. "Auto-encoding variational bayes". In: *arXiv preprint arXiv:1312.6114* (2013).

[125] Aaron Klein and Frank Hutter. "Tabular benchmarks for joint architecture and hyperparameter optimization". In: *arXiv preprint arXiv:1905.04970* (2019).

[126] Steven Kleinegesse and Michael U Gutmann. "Efficient Bayesian experimental design for implicit models". In: *The 22nd International Conference on Artificial Intelligence and Statistics*. PMLR. 2019, pp. 476–485.

[127] Steven Kleinegesse and Michael U Gutmann. "Bayesian experimental design for implicit models by mutual information neural estimation". In: *International Conference on Machine Learning*. PMLR. 2020, pp. 5316–5326.

[128] Agustinus Kristiadi, Alexander Immer, Runa Eschenhagen, and Vincent Fortuin. "Promises and Pitfalls of the Linearized Laplace in Bayesian Optimization". In: *arXiv preprint arXiv:2304.08309* (2023).

[129] Alex Krizhevsky, Geoffrey Hinton, et al. "Learning multiple layers of features from tiny images". In: (2009).

[130] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. "Imagenet classification with deep convolutional neural networks". In: *Advances in neural information processing systems* 25 (2012).

[131] Tze Leung Lai, Herbert Robbins, et al. "Asymptotically efficient adaptive allocation rules". In: *Advances in applied mathematics* 6.1 (1985), pp. 4–22.

[132] Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. "Simple and scalable predictive uncertainty estimation using deep ensembles". In: *Advances in neural information processing systems* 30 (2017).

[133] Rémi Lam, Matthias Poloczek, Peter Frazier, and Karen E Willcox. "Advances in Bayesian optimization with applications in aerospace engineering". In: *2018 AIAA Non-Deterministic Approaches Conference*. 2018, p. 1656.

[134] Pierre Simon Laplace. *Théorie analytique des probabilités*. Courcier, 1814.

[135] Harri Lappalainen and Antti Honkela. "Bayesian non-linear independent component analysis by multi-layer perceptrons". In: *Advances in independent component analysis* (2000), pp. 93–121.

[136] Miguel Lázaro-Gredilla and Anibal Figueiras-Vidal. "Interdomain Gaussian processes for sparse inference using inducing features". In: *Advances in Neural Information Processing Systems* 22 (2009).

[137] Quoc Le, Tamás Sarlós, Alex Smola, et al. "Fastfood-approximating kernel expansions in loglinear time". In: *Proceedings of the international conference on machine learning*. Vol. 85. 2013.

[138] Yann A LeCun, Léon Bottou, Genevieve B Orr, and Klaus-Robert Müller. "Efficient backprop". In: *Neural Networks: Tricks of the Trade*. Springer, 2012, pp. 9–48.

[139] Jaehoon Lee, Yasaman Bahri, Roman Novak, Samuel S Schoenholz, Jeffrey Pennington, and Jascha Sohl-Dickstein. "Deep neural networks as gaussian processes". In: *arXiv preprint arXiv:1711.00165* (2017).

[140] Chunyuan Li, Hao Liu, Changyou Chen, Yuchen Pu, Liqun Chen, Ricardo Henao, and Lawrence Carin. "Alice: Towards understanding adversarial learning for joint distribution matching". In: *Advances in neural information processing systems* 30 (2017).

[141] Yucen Lily Li, Tim GJ Rudner, and Andrew Gordon Wilson. "A Study of Bayesian Neural Network Surrogates for Bayesian Optimization". In: *arXiv preprint arXiv:2305.20028* (2023).

[142] Friedrich Liese and Igor Vajda. "On divergences and informations in statistics and information theory". In: *IEEE Transactions on Information Theory* 52.10 (2006), pp. 4394–4412.

[143] Timothy P Lillicrap, Jonathan J Hunt, Alexander Pritzel, Nicolas Heess, Tom Erez, Yuval Tassa, David Silver, and Daan Wierstra. "Continuous control with deep reinforcement learning". In: *arXiv preprint arXiv:1509.02971* (2015).

[144] Alexander Liniger, Alexander Domahidi, and Manfred Morari. "Optimization-based autonomous racing of 1: 43 scale RC cars". In: *Optimal Control Applications and Methods* 36.5 (2015), pp. 628–647.

[145] Thomas Lipp and Stephen Boyd. "Minimum-time speed optimisation over a fixed path". In: *International Journal of Control* 87.6 (2014), pp. 1297–1311.

[146] Dong C Liu and Jorge Nocedal. "On the limited memory BFGS method for large scale optimization". In: *Mathematical Programming* 45.1-3 (1989), pp. 503–528.

[147] Qiang Liu and Dilin Wang. "Stein variational gradient descent: A general purpose bayesian inference algorithm". In: *Advances in neural information processing systems* 29 (2016).

[148] Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. "Deep learning face attributes in the wild". In: *Proceedings of the IEEE international conference on computer vision*. 2015, pp. 3730–3738.

[149] Wenlong Lyu, Pan Xue, Fan Yang, Changhao Yan, Zhiliang Hong, Xuan Zeng, and Dian Zhou. "An efficient bayesian optimization approach for automated optimization of analog circuits". In: *IEEE Transactions on Circuits and Systems I: Regular Papers* 65.6 (2017), pp. 1954–1967.

[150] David JC MacKay. "A practical Bayesian framework for back-propagation networks". In: *Neural computation* 4.3 (1992), pp. 448–472.

[151] David JC MacKay. "The evidence framework applied to classification networks". In: *Neural computation* 4.5 (1992), pp. 720–736.

[152] David JC MacKay et al. "Introduction to Gaussian processes". In: *NATO ASI series F computer and systems sciences* 168 (1998), pp. 133–166.

[153] David JC MacKay. *Information theory, inference and learning algorithms*. Cambridge university press, 2003.

[154] Chris J Maddison, Andriy Mnih, and Yee Whye Teh. "The concrete distribution: A continuous relaxation of discrete random variables". In: *arXiv preprint arXiv:1611.00712* (2016).

[155] Anastasia Makarova, Ilnura Usmanova, Ilija Bogunovic, and Andreas Krause. "Risk-averse heteroscedastic bayesian optimization". In: *Advances in Neural Information Processing Systems* 34 (2021), pp. 17235–17245.

[156] Alireza Makhzani, Jonathon Shlens, Navdeep Jaitly, Ian Goodfellow, and Brendan Frey. "Adversarial autoencoders". In: *arXiv preprint arXiv:1511.05644* (2015).

[157] Anton Mallasto and Aasa Feragen. "Learning from uncertain curves: The 2-Wasserstein metric for Gaussian processes". In: *Advances in Neural Information Processing Systems* 30 (2017).

[158] Xudong Mao, Qing Li, Haoran Xie, Raymond YK Lau, Zhen Wang, and Stephen Paul Smolley. "Least squares generative adversarial networks". In: *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 2794–2802.

[159] Roman Marchant and Fabio Ramos. "Bayesian optimisation for intelligent environmental monitoring". In: *2012 IEEE/RSJ international conference on intelligent robots and systems*. IEEE. 2012, pp. 2242–2249.

[160]   Bertil Matérn. "Spatial variation: Stachastic models and their application to some problems in forest surveys and other sampling investigations". PhD thesis. Stockholm University, 1960.

[161]   Alexander G de G Matthews, James Hensman, Richard Turner, and Zoubin Ghahramani. "On sparse variational methods and the Kullback-Leibler divergence between stochastic processes". In: *Artificial Intelligence and Statistics*. PMLR. 2016, pp. 231–239.

[162]   Alexander G de G Matthews, Mark Rowland, Jiri Hron, Richard E Turner, and Zoubin Ghahramani. "Gaussian process behaviour in wide deep neural networks". In: *arXiv preprint arXiv:1804.11271* (2018).

[163]   Alexander G. de G. Matthews, Mark van der Wilk, Tom Nickson, Keisuke. Fujii, Alexis Boukouvalas, Pablo León-Villagrá, Zoubin Ghahramani, and James Hensman. "GPflow: A Gaussian process library using TensorFlow". In: *Journal of Machine Learning Research* 18.40 (Apr. 2017), pp. 1–6. URL: http://jmlr.org/papers/v18/16-537.html.

[164]   Alexander Graeme de Garis Matthews. "Scalable Gaussian process inference using variational methods". PhD thesis. University of Cambridge, 2017.

[165]   Warren S McCulloch and Walter Pitts. "A logical calculus of the ideas immanent in nervous activity". In: *The bulletin of mathematical biophysics* 5 (1943), pp. 115–133.

[166]   Aditya Menon and Cheng Soon Ong. "Linking losses for density ratio and class-probability estimation". In: *International Conference on Machine Learning*. 2016, pp. 304–313.

[167]   Lars Mescheder, Sebastian Nowozin, and Andreas Geiger. "Adversarial variational bayes: Unifying variational autoencoders and generative adversarial networks". In: *International conference on machine learning*. PMLR. 2017, pp. 2391–2400.

[168]   Thomas P Minka. "Expectation propagation for approximate Bayesian inference". In: *arXiv preprint arXiv:1301.2294* (2013).

[169]   Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra, and Martin Riedmiller. "Playing atari with deep reinforcement learning". In: *arXiv preprint arXiv:1312.5602* (2013).

[170]   Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fidjeland, Georg Ostrovski, et al. "Human-level control through deep reinforcement learning". In: *nature* 518.7540 (2015), pp. 529–533.

[171]   Jonas Mockus, Vytautas Tiesis, and Antanas Zilinskas. "The application of Bayesian methods for seeking the extremum". In: *Towards Global Optimization* 2.117-129 (1978), p. 2.

[172]   Shakir Mohamed and Balaji Lakshminarayanan. "Learning in implicit generative models". In: *arXiv preprint arXiv:1610.03483* (2016).

[173]   Henry B Moss, Daniel Beck, Javier González, David S Leslie, and Paul Rayson. "BOSS: Bayesian Optimization over String Spaces". In: *arXiv preprint arXiv:2010.00979* (2020).

[174]   Henry B Moss, Sebastian W Ober, and Victor Picheny. "Inducing point allocation for sparse gaussian processes in high-throughput bayesian optimisation". In: *International Conference on Artificial Intelligence and Statistics*. PMLR. 2023, pp. 5213–5230.

[175]   Marina Munkhoeva, Yermek Kapushev, Evgeny Burnaev, and Ivan Oseledets. "Quadrature-based features for kernel approximation". In: *arXiv preprint arXiv:1802.03832* (2018).

[176]   Kevin P Murphy. *Machine learning: a probabilistic perspective*. MIT press, 2012.

[177]   Iain Murray and Ryan P Adams. "Slice sampling covariance hyperparameters of latent Gaussian models". In: *Advances in neural information processing systems* 23 (2010).

[178]   Mojmir Mutny and Andreas Krause. "Efficient high dimensional bayesian optimization with additivity and quadrature fourier features". In: *Advances in Neural Information Processing Systems* 31 (2018).

[179]   Mojmír Mutnỳ and Andreas Krause. "Efficient high dimensional Bayesian optimization with additivity and quadrature Fourier features". In: *Advances in Neural Information Processing Systems* 31 (2019), pp. 9005–9016.

[180]   Radford M Neal. "BAYESIAN LEARNING FOR NEURAL NETWORKS". PhD thesis. University of Toronto, 1995.

[181]   Radford M Neal. "Slice sampling". In: *The annals of statistics* 31.3 (2003), pp. 705–767.

[182]   Radford M Neal et al. "MCMC using Hamiltonian dynamics". In: *Handbook of markov chain monte carlo* 2.11 (2011), p. 2.

[183]   XuanLong Nguyen, Martin J Wainwright, and Michael I Jordan. "Estimating divergence functionals and the likelihood ratio by convex risk minimization". In: *IEEE Transactions on Information Theory* 56.11 (2010), pp. 5847–5861.

[184]   Alexandru Niculescu-Mizil and Rich Caruana. "Predicting good probabilities with supervised learning". In: *Proceedings of the 22nd International Conference on Machine Learning*. 2005, pp. 625–632.

[185]   Sebastian Nowozin, Botond Cseke, and Ryota Tomioka. "f-gan: Training generative neural samplers using variational divergence minimization". In: *Advances in neural information processing systems* 29 (2016).

[186]   Sebastian W Ober, Carl E Rasmussen, and Mark van der Wilk. "The promises and pitfalls of deep kernel learning". In: *Uncertainty in Artificial Intelligence*. PMLR. 2021, pp. 1206–1216.

[187]   Rafael Oliveira, Louis C Tiao, and Fabio T Ramos. "Batch Bayesian Optimisation via Density-Ratio Estimation with Guarantees". In: *Advances in Neural Information Processing Systems*. Ed. by S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh. Vol. 35. Curran Associates, Inc., 2022, pp. 29816–29829. URL: https://bit.ly/oliveira2022batch.

[188]   R OpenAI. "GPT-4 technical report". In: *arXiv* (2023), pp. 2303–08774.

[189]   Manfred Opper and Ole Winther. "Gaussian processes and SVM: Mean field results and leave-one-out". In: (2000).

[190]   Byeong U Park and James S Marron. "Comparison of data-driven bandwidth selectors". In: *Journal of the American Statistical Association* 85.409 (1990), pp. 66–72.

[191]   Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. "Pytorch: An imperative style, high-performance deep learning library". In: *Advances in neural information processing systems* 32 (2019).

[192]   Karl Pearson. "LIII. On lines and planes of closest fit to systems of points in space". In: *The London, Edinburgh, and Dublin philosophical magazine and journal of science* 2.11 (1901), pp. 559–572.

[193]   F. Pedregosa et al. "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.

[194]   Valerio Perrone, Rodolphe Jenatton, Matthias W Seeger, and Cédric Archambeau. "Scalable hyperparameter transfer learning". In: *Advances in neural information processing systems* 31 (2018).

[195]   John Platt et al. "Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods". In: *Advances in Large Margin Classifiers* 10.3 (1999), pp. 61–74.

[196]   Yuchen Pu, Weiyao Wang, Ricardo Henao, Liqun Chen, Zhe Gan, Chunyuan Li, and Lawrence Carin. "Adversarial symmetric variational autoencoder". In: *Advances in neural information processing systems* 30 (2017).

[197]   Jing Qin. "Inferences for case-control and semiparametric two-sample density ratio models". In: *Biometrika* 85.3 (1998), pp. 619–630.

[198]   Joaquin Quinonero-Candela and Carl Edward Rasmussen. "A unifying view of sparse approximate Gaussian process regression". In: *The Journal of Machine Learning Research* 6 (2005), pp. 1939–1959.

[199]   Jack W Rae, Sebastian Borgeaud, Trevor Cai, Katie Millican, Jordan Hoffmann, Francis Song, John Aslanides, Sarah Henderson, Roman Ring, Susannah Young, et al. "Scaling language models: Methods, analysis & insights from training gopher". In: *arXiv preprint arXiv:2112.11446* (2021).

[200]   Ali Rahimi and Benjamin Recht. "Random Features for Large-Scale Kernel Machines." In: *NIPS*. Vol. 3. 4. Citeseer. 2007, p. 5.

[201]   Ali Rahimi and Benjamin Recht. "Weighted Sums of Random Kitchen Sinks: Replacing minimization with randomization in learning". In: *Advances in Neural Information Processing Systems* 21 (2008), pp. 1313–1320.

[202]   Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. "Hierarchical text-conditional image generation with clip latents". In: *arXiv preprint arXiv:2204.06125* 1.2 (2022), p. 3.

[203]   Rajesh Ranganath, Sean Gerrish, and David Blei. "Black box variational inference". In: *Artificial intelligence and statistics*. PMLR. 2014, pp. 814–822.

[204]   Esteban Real, Alok Aggarwal, Yanping Huang, and Quoc V Le. "Regularized evolution for image classifier architecture search". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 33. 2019, pp. 4780–4789.

[205]   Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. "You only look once: Unified, real-time object detection". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 779–788.

[206]   Danilo Jimenez Rezende, Shakir Mohamed, and Daan Wierstra. "Stochastic backpropagation and approximate inference in deep generative models". In: *International conference on machine learning*. PMLR. 2014, pp. 1278–1286.

[207]   R Tyrrell Rockafellar. *Convex analysis*. Vol. 11. Princeton university press, 1997.

[208]   Joseph Lee Rodgers, W Alan Nicewander, and Larry Toothaker. "Linearly independent, orthogonal, and uncorrelated variables". In: *The American Statistician* 38.2 (1984), pp. 133–134.

[209] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. "High-resolution image synthesis with latent diffusion models". In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2022, pp. 10684–10695.

[210] Philip A Romero, Andreas Krause, and Frances H Arnold. "Navigating the protein fitness landscape with Gaussian processes". In: *Proceedings of the National Academy of Sciences* 110.3 (2013), E193–E201.

[211] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation". In: *Medical Image Computing and Computer-Assisted Intervention– MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*. Springer. 2015, pp. 234–241.

[212] Lorenzo Rosasco, Ernesto De Vito, Andrea Caponnetto, Michele Piana, and Alessandro Verri. "Are loss functions all the same?" In: *Neural computation* 16.5 (2004), pp. 1063–1076.

[213] Frank Rosenblatt. "The perceptron: a probabilistic model for information storage and organization in the brain." In: *Psychological review* 65.6 (1958), p. 386.

[214] Ugo Rosolia and Francesco Borrelli. "Learning how to autonomously race a car: a predictive control approach". In: *IEEE Transactions on Control Systems Technology* 28.6 (2019), pp. 2713–2719.

[215] Sam Roweis and Zoubin Ghahramani. "A unifying review of linear Gaussian models". In: *Neural computation* 11.2 (1999), pp. 305–345.

[216] Amir Saffari, Christian Leistner, Jakob Santner, Martin Godec, and Horst Bischof. "On-line random forests". In: *2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops*. IEEE. 2009, pp. 1393–1400.

[217] Hugh Salimbeni, Ching-An Cheng, Byron Boots, and Marc Deisenroth. "Orthogonally decoupled variational Gaussian processes". In: *Advances in Neural Information Processing Systems* 31 (2018).

[218] David Salinas, Matthias Seeger, Aaron Klein, Valerio Perrone, Martin Wistuba, and Cedric Archambeau. "Syne tune: A library for large scale hyperparameter tuning and reproducible research". In: *International Conference on Automated Machine Learning*. PMLR. 2022, pp. 16–1.

[219] Warren Scott, Peter Frazier, and Warren Powell. "The correlated knowledge gradient for simulation optimization of continuous parameters using gaussian process regression". In: *SIAM Journal on Optimization* 21.3 (2011), pp. 996–1026.

[220] Matthias W Seeger, Christopher KI Williams, and Neil D Lawrence. "Fast forward selection to speed up sparse Gaussian process regression". In: *International Workshop on Artificial Intelligence and Statistics*. PMLR. 2003, pp. 254–261.

[221] Atsuto Seko, Atsushi Togo, Hiroyuki Hayashi, Koji Tsuda, Laurent Chaput, and Isao Tanaka. "Prediction of low-thermal-conductivity compounds with first-principles anharmonic lattice-dynamics calculations and Bayesian optimization". In: *Physical review letters* 115.20 (2015), p. 205901.

[222] Bobak Shahriari, Kevin Swersky, Ziyu Wang, Ryan P Adams, and Nando De Freitas. "Taking the human out of the loop: A review of Bayesian optimization". In: *Proceedings of the IEEE* 104.1 (2015), pp. 148–175.

[223] Simon J Sheather and Michael C Jones. "A reliable data-based bandwidth selection method for kernel density estimation". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 53.3 (1991), pp. 683–690.

[224] Jiaxin Shi, Michalis Titsias, and Andriy Mnih. "Sparse orthogonal variational inference for Gaussian processes". In: *International Conference on Artificial Intelligence and Statistics*. PMLR. 2020, pp. 1932–1942.

[225] Mohammad Shoeybi, Mostofa Patwary, Raul Puri, Patrick LeGresley, Jared Casper, and Bryan Catanzaro. "Megatron-lm: Training multi-billion parameter language models using model parallelism". In: *arXiv preprint arXiv:1909.08053* (2019).

[226] David Silver, Aja Huang, Chris J Maddison, Arthur Guez, Laurent Sifre, George Van Den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, et al. "Mastering the game of Go with deep neural networks and tree search". In: *nature* 529.7587 (2016), pp. 484–489.

[227] Bernard W Silverman. *Density estimation for statistics and data analysis*. Vol. 26. CRC Press, 1986.

[228] Edward Snelson and Zoubin Ghahramani. "Sparse Gaussian processes using pseudo-inputs". In: *Advances in neural information processing systems* 18 (2005).

[229] Edward Snelson and Zoubin Ghahramani. "Local and global sparse Gaussian process approximations". In: *Artificial Intelligence and Statistics*. PMLR. 2007, pp. 524–531.

[230]   Jasper Snoek, Hugo Larochelle, and Ryan P Adams. "Practical Bayesian optimization of machine learning algorithms". In: *Advances in Neural Information Processing Systems* 25 (2012), pp. 2951–2959.

[231]   Jasper Snoek, Oren Rippel, Kevin Swersky, Ryan Kiros, Nadathur Satish, Narayanan Sundaram, Mostofa Patwary, Mr Prabhat, and Ryan Adams. "Scalable bayesian optimization using deep neural networks". In: *International conference on machine learning*. PMLR. 2015, pp. 2171–2180.

[232]   Jasper Snoek, Kevin Swersky, Rich Zemel, and Ryan Adams. "Input warping for Bayesian optimization of non-stationary functions". In: *International Conference on Machine Learning*. PMLR. 2014, pp. 1674–1682.

[233]   Casper Kaae Sønderby, Jose Caballero, Lucas Theis, Wenzhe Shi, and Ferenc Huszár. "Amortised map inference for image super-resolution". In: *arXiv preprint arXiv:1610.04490* (2016).

[234]   Jiaming Song and Stefano Ermon. "Likelihood-free Density Ratio Acquisition Functions are not Equivalent to Expected Improvements". In: *Bayesian Deep Learning Workshop at NeurIPS*. 2021.

[235]   Jiaming Song, Lantao Yu, Willie Neiswanger, and Stefano Ermon. "A general recipe for likelihood-free Bayesian optimization". In: *International Conference on Machine Learning*. PMLR. 2022, pp. 20384–20404.

[236]   C Spearman. "" General Intelligence," Objectively Determined and Measured". In: *The American Journal of Psychology* 15.2 (1904), pp. 201–292.

[237]   Jost Tobias Springenberg, Aaron Klein, Stefan Falkner, and Frank Hutter. "Bayesian optimization with robust Bayesian neural networks". In: *Advances in neural information processing systems* 29 (2016).

[238]   Niranjan Srinivas, Andreas Krause, Sham M Kakade, and Matthias Seeger. "Gaussian process optimization in the bandit setting: No regret and experimental design". In: *arXiv preprint arXiv:0912.3995* (2009).

[239]   Akash Srivastava, Lazar Valkov, Chris Russell, Michael U Gutmann, and Charles Sutton. "Veegan: Reducing mode collapse in gans using implicit variational learning". In: *Advances in neural information processing systems* 30 (2017).

[240]   Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. "Dropout: a simple way to prevent neural networks from overfitting". In: *The journal of machine learning research* 15.1 (2014), pp. 1929–1958.

[241]   Michael L Stein. *Interpolation of spatial data: some theory for kriging*. Springer Science & Business Media, 1999.

[242]   Josef Stoer and Roland Bulirsch. *Introduction to numerical analysis*. Vol. 12. Springer Science & Business Media, 2013.

[243]   R. Storn and K. Price. "Differential Evolution – A Simple and Efficient Heuristic for global Optimization over Continuous Spaces". In: *Journal of Global Optimization* (1997).

[244]   Masashi Sugiyama, Shinichi Nakajima, Hisashi Kashima, Paul V Buenau, and Motoaki Kawanabe. "Direct importance estimation with model selection and its application to covariate shift adaptation". In: *Advances in Neural Information Processing Systems*. 2008, pp. 1433–1440.

[245]   Masashi Sugiyama, Taiji Suzuki, and Takafumi Kanamori. *Density Ratio Estimation in Machine Learning*. Cambridge University Press, 2012.

[246]   Shengyang Sun, Jiaxin Shi, and Roger Baker Grosse. "Neural Networks as Inter-Domain Inducing Points". In: *Third Symposium on Advances in Approximate Bayesian Inference*. 2020.

[247]   Shengyang Sun, Jiaxin Shi, Andrew Gordon Gordon Wilson, and Roger B Grosse. "Scalable Variational Gaussian Processes via Harmonic Kernel Decomposition". In: *International Conference on Machine Learning*. PMLR. 2021, pp. 9955–9965.

[248]   Dougal J Sutherland and Jeff Schneider. "On the error of random Fourier features". In: *Proceedings of the Thirty-First Conference on Uncertainty in Artificial Intelligence*. 2015, pp. 862–871.

[249]   Kevin Swersky, Jasper Snoek, and Ryan P Adams. "Multi-task Bayesian optimization". In: *Advances in Neural Information Processing Systems*. 2013, pp. 2004–2012.

[250]   George R Terrell and David W Scott. "Variable kernel density estimation". In: *The Annals of Statistics* (1992), pp. 1236–1265.

[251]   Owen Thomas, Ritabrata Dutta, Jukka Corander, Samuel Kaski, and Michael U Gutmann. "Likelihood-free inference by ratio estimation". In: *Bayesian Analysis* 17.1 (2022), pp. 1–31.

[252]   William R Thompson. "On the likelihood that one unknown probability exceeds another in view of the evidence of two samples". In: *Biometrika* 25.3-4 (1933), pp. 285–294.

[253]   Louis C Tiao, Aaron Klein, Matthias W Seeger, Edwin V Bonilla, Cedric Archambeau, and Fabio T Ramos. "BORE: Bayesian Optimization by Density-Ratio Estimation". In: *Proceedings of the 38th International Conference on Machine Learning*. Ed. by Marina Meila and Tong Zhang. Vol. 139. Proceedings of Machine Learning Research. PMLR, July 2021, pp. 10289–10300. URL:

https://proceedings.mlr.press/v139/tiao21a.html. **Accepted as *Oral* presentation.**

[254]   Robert Tibshirani. "Regression shrinkage and selection via the lasso". In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 58.1 (1996), pp. 267–288.

[255]   Michael E Tipping and Christopher M Bishop. "Probabilistic principal component analysis". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 61.3 (1999), pp. 611–622.

[256]   Michalis Titsias. "Variational learning of inducing variables in sparse Gaussian processes". In: *Artificial Intelligence and Statistics*. PMLR. 2009, pp. 567–574.

[257]   Michalis K Titsias. "Variational model selection for sparse Gaussian process regression". In: *Report, University of Manchester, UK* (2009).

[258]   Michalis K Titsias. "Learning model reparametrizations: Implicit variational inference by fitting MCMC distributions". In: *arXiv preprint arXiv:1708.01529* (2017).

[259]   Hakki Mert Torun, Madhavan Swaminathan, Anto Kavungal Davis, and Mohamed Lamine Faycal Bellaredj. "A global Bayesian optimization algorithm and its application to integrated system design". In: *IEEE Transactions on Very Large Scale Integration (VLSI) Systems* 26.4 (2018), pp. 792–802.

[260]   Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, et al. "Llama 2: Open foundation and fine-tuned chat models". In: *arXiv preprint arXiv:2307.09288* (2023).

[261]   Dustin Tran, Rajesh Ranganath, and David Blei. "Hierarchical implicit models and likelihood-free variational inference". In: *Advances in Neural Information Processing Systems* 30 (2017).

[262]   Yuta Tsuboi, Hisashi Kashima, Shohei Hido, Steffen Bickel, and Masashi Sugiyama. "Direct density ratio estimation for large-scale covariate shift adaptation". In: *Journal of Information Processing* 17 (2009), pp. 138–155.

[263]   Zhuowen Tu. "Learning generative models via discriminative approaches". In: *2007 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE. 2007, pp. 1–8.

[264]   Ryan Turner, David Eriksson, Michael McCourt, Juha Kiili, Eero Laaksonen, Zhen Xu, and Isabelle Guyon. "Bayesian optimization is superior to random search for machine learning hyperparameter tuning: Analysis of the black-box optimization challenge 2020". In: *NeurIPS 2020 Competition and Demonstration Track*. PMLR. 2021, pp. 3–26.

[265] Masatoshi Uehara, Issei Sato, Masahiro Suzuki, Kotaro Nakayama, and Yutaka Matsuo. "Generative adversarial nets from a density ratio estimation perspective". In: *arXiv preprint arXiv:1610.02920* (2016).

[266] Mark Van der Wilk. "Sparse Gaussian process approximations and applications". PhD thesis. University of Cambridge, 2019.

[267] Vladimir Vapnik. *The nature of statistical learning theory*. Springer Science & Business Media, 2013.

[268] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. "Attention is all you need". In: *Advances in neural information processing systems* 30 (2017).

[269] Pauli Virtanen et al. "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python". In: *Nature Methods* 17 (2020), pp. 261–272. DOI: 10.1038/s41592-019-0686-2.

[270] Martin J Wainwright, Michael I Jordan, et al. "Graphical models, exponential families, and variational inference". In: *Foundations and Trends® in Machine Learning* 1.1–2 (2008), pp. 1–305.

[271] Ke Wang, Geoff Pleiss, Jacob Gardner, Stephen Tyree, Kilian Q Weinberger, and Andrew Gordon Wilson. "Exact Gaussian processes on a million data points". In: *Advances in neural information processing systems* 32 (2019).

[272] Zi Wang, Clement Gehring, Pushmeet Kohli, and Stefanie Jegelka. "Batched large-scale Bayesian optimization in high-dimensional spaces". In: *International Conference on Artificial Intelligence and Statistics*. PMLR. 2018, pp. 745–754.

[273] Zi Wang and Stefanie Jegelka. "Max-value entropy search for efficient Bayesian optimization". In: *arXiv preprint arXiv:1703.01968* (2017).

[274] Florian Wenzel, Jasper Snoek, Dustin Tran, and Rodolphe Jenatton. "Hyperparameter ensembles for robustness and uncertainty quantification". In: *arXiv preprint arXiv:2006.13570* (2020).

[275] Colin White, Willie Neiswanger, and Yash Savani. "Bananas: Bayesian optimization with neural architectures for neural architecture search". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 35. 12. 2021, pp. 10293–10301.

[276] Paul B Wigley, Patrick J Everitt, Anton van den Hengel, John W Bastian, Mahasen A Sooriyabandara, Gordon D McDonald, Kyle S Hardman, Ciaron D Quinlivan, P Manju, Carlos CN Kuhn, et al. "Fast machine-learning online optimization of ultra-cold-atom experiments". In: *Scientific reports* 6.1 (2016), p. 25890.

[277] Mark van der Wilk, Vincent Dutordoir, ST John, Artem Arte-
       mev, Vincent Adam, and James Hensman. "A framework for
       interdomain and multioutput Gaussian processes". In: *arXiv
       preprint arXiv:2003.01115* (2020).

[278] Christopher KI Williams and David Barber. "Bayesian classifica-
       tion with Gaussian processes". In: *IEEE Transactions on Pattern
       Analysis and Machine Intelligence* 20.12 (1998), pp. 1342–1351.

[279] Christopher KI Williams and Carl Edward Rasmussen. *Gaussian
       processes for machine learning*. Vol. 2. 3. MIT press Cambridge,
       MA, 2006.

[280] Christopher KI Williams and Matthias Seeger. "Using the Nys-
       tröm method to speed up kernel machines". In: *Proceedings of
       the 13th International Conference on Neural Information Processing
       Systems*. 2000, pp. 661–667.

[281] Andrew Gordon Wilson, Zhiting Hu, Ruslan Salakhutdinov,
       and Eric P Xing. "Deep kernel learning". In: *Artificial intelligence
       and statistics*. PMLR. 2016, pp. 370–378.

[282] James Wilson, Viacheslav Borovitskiy, Alexander Terenin, Pe-
       ter Mostowsky, and Marc Deisenroth. "Efficiently sampling
       functions from Gaussian process posteriors". In: *International
       Conference on Machine Learning*. PMLR. 2020, pp. 10292–10302.

[283] James Wilson, Frank Hutter, and Marc Deisenroth. "Maxi-
       mizing acquisition functions for Bayesian optimization". In:
       Advances in Neural Information Processing Systems 31 (2018),
       pp. 9884–9895.

[284] James T Wilson, Viacheslav Borovitskiy, Alexander Terenin,
       Peter Mostowsky, and Marc Peter Deisenroth. "Pathwise con-
       ditioning of gaussian processes". In: *The Journal of Machine
       Learning Research* 22.1 (2021), pp. 4741–4787.

[285] Makoto Yamada, Taiji Suzuki, Takafumi Kanamori, Hirotaka
       Hachiya, and Masashi Sugiyama. "Relative density-ratio es-
       timation for robust distribution comparison". In: *Advances in
       Neural Information Processing Systems*. 2011, pp. 594–602.

[286] Jiyan Yang, Vikas Sindhwani, Haim Avron, and Michael Ma-
       honey. "Quasi-Monte Carlo feature maps for shift-invariant
       kernels". In: *International Conference on Machine Learning*. PMLR.
       2014, pp. 485–493.

[287] Kevin K Yang, Zachary Wu, and Frances H Arnold. "Machine-
       learning-guided directed evolution for protein engineering".
       In: *Nature methods* 16.8 (2019), pp. 687–694.

[288] Tianbao Yang, Yu-Feng Li, Mehrdad Mahdavi, Rong Jin, and
       Zhi-Hua Zhou. "Nyström method vs random fourier features:
       A theoretical and empirical comparison". In: *Advances in neural
       information processing systems* 25 (2012), pp. 476–484.

[289] Zichao Yang, Andrew Wilson, Alex Smola, and Le Song. "A la carte–learning fast kernels". In: *Artificial Intelligence and Statistics*. PMLR. 2015, pp. 1098–1106.

[290] Felix Xinnan X Yu, Ananda Theertha Suresh, Krzysztof M Choromanski, Daniel N Holtmann-Rice, and Sanjiv Kumar. "Orthogonal random features". In: *Advances in neural information processing systems* 29 (2016), pp. 1975–1983.

[291] Bianca Zadrozny and Charles Elkan. "Obtaining calibrated probability estimates from decision trees and naive Bayesian classifiers". In: *ICML*. Vol. 1. Citeseer. 2001, pp. 609–616.

[292] Bianca Zadrozny and Charles Elkan. "Transforming classifier scores into accurate multiclass probability estimates". In: *Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2002, pp. 694–699.

[293] Ciyou Zhu, Richard H Byrd, Peihuang Lu, and Jorge Nocedal. "Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization". In: *ACM Transactions on mathematical software (TOMS)* 23.4 (1997), pp. 550–560.

[294] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. "Unpaired image-to-image translation using cycle-consistent adversarial networks". In: *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 2223–2232.