The following is a list of the references I mentioned throughout the lecture along with some additional reference material which may help you get a broader view of the literature.

Lecture 1

1. Bottou, L. and Cun, Y., 2003. Large scale online learning. *Advances in neural information processing systems*, *16*.
2. Marchenko, V.A. and Pastur, L.A., 1967. Distribution of eigenvalues for some sets of random matrices. *Mat. Sb.(NS)*, *72*(114), p.4.
3. Baik, J., Ben Arous, G. and Péché, S., 2005. Phase Transition of the Largest Eigenvalue for Nonnull Complex Sample Covariance Matrices. *Annals of Probability*, pp.1643-1697.
4. Oja, E., 1982. Simplified neuron model as a principal component analyzer. *Journal of mathematical biology*, *15*, pp.267-273.
5. ImageNET https://image-net.org/about.php
6. CIFAR 10 https://www.cs.toronto.edu/~kriz/cifar.html
7. Ben Arous, G., Gheissari, R. and Jagannath, A., 2021. Online stochastic gradient descent on non-convex losses from high-dimensional inference. *Journal of Machine Learning Research*, *22*(106), pp.1-51.

Here are some references for single and multi-index models. This list includes some additional material to help you get a better grasp of the field but is by no means exhaustive.

8. Damian, A., Pillaud-Vivien, L., Lee, J. and Bruna, J., 2024, June. Computational-Statistical Gaps in Gaussian Single-Index Models. In *The Thirty Seventh Annual Conference on Learning Theory* (pp. 1262-1262). PMLR.
9. Mondelli, M. and Montanari, A., 2018, July. Fundamental limits of weak recovery with applications to phase retrieval. In *Conference On Learning Theory* (pp. 1445-1450). PMLR.
10. Lu, Y.M. and Li, G., 2020. Phase transitions of spectral initialization for high-dimensional non-convex estimation. *Information and Inference: A Journal of the IMA*, *9*(3), pp.507-541.
11. Tan, Y.S. and Vershynin, R., 2023. Online stochastic gradient descent with arbitrary initialization solves non-smooth, non-convex phase retrieval. *Journal of Machine Learning Research*, *24*(58), pp.1-47.
12. Damian, A., Nichani, E., Ge, R. and Lee, J.D., 2023. Smoothing the landscape boosts the signal for sgd: Optimal sample complexity for learning single index models. *Advances in Neural Information Processing Systems*, *36*, pp.752-784.
13. Abbe, E., Adsera, E.B. and Misiakiewicz, T., 2023, July. Sgd learning on neural networks: leap complexity and saddle-to-saddle dynamics. In *The Thirty Sixth Annual Conference on Learning Theory* (pp. 2552-2623). PMLR.
14. Ba, J., Erdogdu, M.A., Suzuki, T., Wang, Z. and Wu, D., 2023. Learning in the presence of low-dimensional structure: a spiked random matrix perspective. *Advances in Neural Information Processing Systems*, *36*, pp.17420-17449.
15. Dandi, Y., Troiani, E., Arnaboldi, L., Pesce, L., Zdeborova, L. and Krzakala, F., The Benefits of Reusing Batches for Gradient Descent in Two-Layer Networks: Breaking the Curse of Information and Leap Exponents. In *Forty-first International Conference on Machine Learning*.
16. Lee, J.D., Oko, K., Suzuki, T. and Wu, D., 2024. Neural network learns low-dimensional polynomials with sgd near the information-theoretic limit. *Advances in Neural Information Processing Systems*, *37*, pp.58716-58756.

Lecture 2:

Classical asymptotics

1. Robbins, H. and Monro, S., 1951. A stochastic approximation method. *The annals of mathematical statistics*, pp.400-407.
2. McLeish, D.L., 1976. Functional and random central limit theorems for the Robbins-Munro process. *Journal of Applied Probability*, *13*(1), pp.148-154.
3. Dupuis, P. and Kushner, H.J., 1989. Stochastic approximation and large deviations: Upper bounds and wp 1 convergence. *SIAM Journal on Control and Optimization*, *27*(5), pp.1108-1135.
4. Stroock, D.W. and Varadhan, S.S., 2007. *Multidimensional diffusion processes*. Springer.

Infinite width/Hydrodynamic scaling

5. Mei, S., Montanari, A. and Nguyen, P.M., 2018. A mean field view of the landscape of two-layer neural networks. *Proceedings of the National Academy of Sciences*, *115*(33), pp.E7665-E7671.
6. Rotskoff, G. and Vanden-Eijnden, E., 2022. Trainability and accuracy of artificial neural networks: An interacting particle system approach. *Communications on Pure and Applied Mathematics*, *75*(9), pp.1889-1935.

High-dimensional scaling

7. Saad, D. and Solla, S.A., 1995. Exact solution for on-line learning in multilayer neural networks. *Physical Review Letters*, *74*(21), p.4337.
8. Goldt, S., Advani, M., Saxe, A.M., Krzakala, F. and Zdeborová, L., 2019. Dynamics of stochastic gradient descent for two-layer neural networks in the teacher-student setup. *Advances in neural information processing systems*, *32*.
9. Refinetti, M., Goldt, S., Krzakala, F. and Zdeborová, L., 2021, July. Classifying high-dimensional gaussian mixtures: Where kernel methods fail and neural networks succeed. In *International Conference on Machine Learning* (pp. 8936-8947). PMLR.
10. Ben Arous, G., Gheissari, R. and Jagannath, A., 2024. High-dimensional limit theorems for SGD: Effective dynamics and critical scaling. *Communications on Pure and Applied Mathematics, 77(3), pp.2030-2080*. [Conference version in NeuRIPS 2022]
11. Collins-Woodfin, E., Paquette, C., Paquette, E. and Seroussi, I., 2024. Hitting the high-dimensional notes: An ode for SGD learning dynamics on GLMs and multi-index models. *Information and Inference: A Journal of the IMA*, *13*(4),

Lecture 3

1. Rao, C. R. (1945): Information and the accuracy attainable in the estimation of statistical parameters. *Bull. Calcutta Math. Soc., 37, 81-91.*
2. LeCun, Y., Bottou, L., Orr, G.B. and Müller, K.R., 2002. Efficient backprop. In *Neural networks: Tricks of the trade* (pp. 9-50). Berlin, Heidelberg: Springer Berlin Heidelberg.
3. Sagun, L., Bottou, L. and LeCun, Y., 2016. Eigenvalues of the hessian in deep learning: Singularity and beyond. *arXiv preprint arXiv:1611.07476*.
4. Sagun, L., Evci, U., Guney, V.U., Dauphin, Y. and Bottou, L., 2017. Empirical analysis of the hessian of over-parametrized neural networks. *arXiv preprint arXiv:1706.04454*.
5. Papyan, V., 2019. Measurements of three-level hierarchical structure in the outliers in the spectrum of deepnet hessians. *arXiv preprint arXiv:1901.08244*.
6. Papyan, V., Han, X.Y. and Donoho, D.L., 2020. Prevalence of neural collapse during the terminal phase of deep learning training. *Proceedings of the National Academy of Sciences*, *117*(40), pp.24652-24663
7. Marchenko, V.A. and Pastur, L.A., 1967. Distribution of eigenvalues for some sets of random matrices. *Mat. Sb.(NS)*, *72*(114), p.4.
8. Silverstein, J.W. and Bai, Z.D., 1995. On the empirical distribution of eigenvalues of a class of large dimensional random matrices. *Journal of Multivariate analysis*, *54*(2), pp.175-192.
9. Anderson, G.W., Guionnet, A. and Zeitouni, O., 2010. *An introduction to random matrices* (No. 118). Cambridge university press.
10. Knowles, A. and Yin, J., 2017. Anisotropic local laws for random matrices. *Probability Theory and Related Fields*, *169*, pp.257-352.
11. Erdos, L., 2019. The matrix Dyson equation and its applications for random matrices. *arXiv preprint arXiv:1903.10060*.
12. Ben Arous, G., Gheissari, R., Huang, J. and Jagannath, A., High-dimensional SGD aligns with emerging outlier eigenspaces. In *The Twelfth International Conference on Learning Representations*. (2024)
13. Ben Arous, G., Gheissari, R., Huang, J. and Jagannath, A., 2025. Local geometry of high-dimensional mixture models: Effective spectral theory and dynamical transitions. *arXiv preprint arXiv:2502.15655*.
14. Liao, Z. and Mahoney, M.W., 2021. Hessian eigenspectra of more realistic nonlinear models. *Advances in Neural Information Processing Systems*, *34*, pp.20104-20117.
15. Baskerville, N.P., Keating, J.P., Mezzadri, F., Najnudel, J. and Granziol, D., 2022. Universal characteristics of deep neural network loss surfaces from random matrix theory. *Journal of Physics A: Mathematical and Theoretical*, *55*(49), p.494002.
16. Garrod, C. and Keating, J.P., 2024. Unifying low dimensional observations in deep learning through the deep linear unconstrained feature model. *arXiv preprint arXiv:2404.06106*.
17. Pennington, J. and Worah, P., 2018. The spectrum of the fisher information matrix of a single-hidden-layer neural network. *Advances in neural information processing systems*, *31*.
18. Mondelli, M. and Montanari, A., 2018, July. Fundamental limits of weak recovery with applications to phase retrieval. In *Conference On Learning Theory* (pp. 1445-1450). PMLR.
19. Lu, Y.M. and Li, G., 2020. Phase transitions of spectral initialization for high-dimensional non-convex estimation. *Information and Inference: A Journal of the IMA*, *9*(3), pp.507-541.