

Early History of Support Vector Machines

- In 1936, R. A. Fisher suggested the first algorithm for pattern recognition (Fisher 1936).
- Aronszajn (1950) introduced the ‘Theory of Reproducing Kernels’.
- In 1957 Frank Rosenblatt invented a linear classifier called the perceptron (the simplest kind of feedforward neural network), see Rosenblatt (1962).
- Vapnik and Lerner (1963) introduce the Generalized Portrait algorithm (the algorithm implemented by support vector machines is a nonlinear generalization of the Generalized Portrait algorithm).
- Aizerman, Braverman and Rozonoer (1964) introduced the geometrical interpretation of the kernels as inner products in a feature space.
- Vapnik and Chervonenkis (1964) further develop the Generalized Portrait algorithm.
- Cover (1965) discussed large margin hyperplanes in the input space and also sparseness.
- Similar optimisation techniques were used in pattern recognition by Mangasarian (1965).
- The use of slack variables to overcome the problem of noise and nonseparability was introduced by Smith (1968).
- Duda and Hart (1973) discuss large margin hyperplanes in the input space.
- The field of ‘statistical learning theory’ began with Vapnik and Chervonenkis (1974) (in Russian).
- SVMs can be said to have started when statistical learning theory was developed further with Vapnik (1979) (in Russian).
- Vapnik and Tschervonenkis (1979) wrote a German translation of Vapnik and Chervonenkis’s 1974 book.
- Vapnik (1982) wrote an English translation of his 1979 book.
- See also the PhD thesis by Hassoun (1986) for related early work.
- Several statistical mechanics papers (for example Anlauf and Biehl (1989)) suggested using large margin hyperplanes in the input space.
- Poggio and Girosi (1990) and Wahba (1990) discuss the use of kernels.
- Bennett and Mangasarian (1992) improved upon Smith’s 1968 work on slack variables.
- SVMs close to their current form were first introduced with a paper at the COLT 1992 conference (Boser, Guyon and Vapnik 1992).
- In 1995 the soft margin classifier was introduced by Cortes and Vapnik (1995); in the same year the algorithm was extended to the case of regression by Vapnik (1995) in *The Nature of Statistical Learning Theory*.
- The papers by Bartlett (1998) and Shawe-Taylor, *et al.* (1998) gave the first rigorous statistical bound on the generalisation of hard margin SVMs.
- Shawe-Taylor and Cristianini (2000) gave statistical bounds on the generalisation of soft margin algorithms and for the regression case.

Bibliography

- AIZERMAN, M. A., E. M. BRAVERMAN, and L. I. ROZONOER, 1964. Theoretical foundations of the potential function method in pattern recognition learning. *Automation and Remote Control*, **25**, 821–837.
- ANLAUF, J. K., and M. BIEHL, 1989. The adatron: An adaptive perceptron algorithm. *Europhysics Letters*, **10**(7), 687–692.
- ARONSZAJN, N., 1950. Theory of reproducing kernels. *Transactions of the American Mathematical Society*, **68**(3), 337–404.
- BARTLETT, Peter L., 1998. The sample complexity of pattern classification with neural networks: The size of the weights is more important than the size of the network. *IEEE Transactions on Information Theory*, **44**(2), 525–536.

- BENNETT, Kristin P., and O. L. MANGASARIAN, 1992. Robust linear programming discrimination of two linearly inseparable sets. *Optimization Methods and Software*, **1**, 23–34.
- BOSER, Bernhard E., Isabelle M. GUYON, and Vladimir N. VAPNIK, 1992. A training algorithm for optimal margin classifiers. In: *COLT '92: Proceedings of the Fifth Annual Workshop on Computational Learning Theory*. New York, NY, USA: ACM Press, pp. 144–152.
- CORTES, Corinna, and Vladimir VAPNIK, 1995. Support-vector networks. *Machine Learning*, **20**(3), 273–297.
- COVER, Thomas M., 1965. Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition. *IEEE Transactions on Electronic Computers*, **14**(3), 326–334.
- DUDA, Richard O., and Peter E. HART, 1973. *Pattern Classification and Scene Analysis*. New York: John Wiley & Sons Inc.
- FISHER, R. A., 1936. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, **7**, 111–132.
- HASSOUN, M. H., 1986. *Optical Threshold Gates and Logical Signal Processing*. Ph. D. thesis, Wayne State University, Detroit, USA.
- MANGASARIAN, O. L., 1965. Linear and nonlinear separation of patterns by linear programming. *Operations Research*, **13**(3), 444–452.
- POGGIO, Tomaso, and Federico GIROSI, 1990. Networks for approximation and learning. *Proceedings of the IEEE*, **78**(9), 1481–1497.
- ROSENBLATT, Frank, 1962. *Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms*. Washington DC: Spartan Books.
- SHAWE-TAYLOR, John, *et al.*, 1998. Structural risk minimization over datadependent hierarchies. *IEEE Transactions on Information Theory*, **44**(5), 1926–1940.
- SHAWE-TAYLOR, John, and Nello CRISTIANINI, 2000. Margin distribution and soft margin. In: Alexander J. SMOLA, *et al.*, eds. *Advances in Large Margin Classifiers*. The MIT Press, pp. 349–358.
- SMITH, F. W., 1968. Pattern classifier design by linear programming. *IEEE Transactions on Computers*, **C-17**(4), 367–372.
- VAPNIK, V., 1979. *Estimation of Dependences Based on Empirical Data* [in Russian]. Moscow: Nauka.
- VAPNIK, Vladimir, 1982. *Estimation of Dependences Based on Empirical Data*. Springer Verlag.
- VAPNIK, V., and A. CHERVONENKIS, 1964. A note on one class of perceptrons. *Automation and Remote Control*, **25**.
- VAPNIK, V., and A. LERNER, 1963. Pattern recognition using generalized portrait method. *Automation and Remote Control*, **24**, 774–780.
- VAPNIK, Vladimir N., 1995. *The Nature of Statistical Learning Theory*. Springer-Verlag New York, Inc.
- VAPNIK, V. N., and A. Ya. CHERVONENKIS, 1974. *Teoriya raspoznavaniya obrazov: Statisticheskie problemy obucheniya*. (Russian) [Theory of pattern recognition: Statistical problems of learning]. Moscow: Nauka.
- WAHBA, Grace, 1990. *Spline Models for Observational Data*. Volume 59 of CBMS-NSF Regional Conference Series in Applied Mathematics. Philadelphia, PA, USA: SIAM: Society for Industrial and Applied Mathematics.
- WAPNIK, W. N., and A. J. TSCHERWONENKIS, 1979. *Theorie der Zeichenerkennung*. (German) [Theory of pattern recognition]. Berlin: Akademie-Verlag.