thesis notes

julia andronowitz

2022-2023

1 To Do

- add how kernel fits into solution equation
- calibration and regression-like problems
- looking for outliers
- compare R and SAS SVM modules
- e1071 documentation?

done

- add cost and gamma values to figures
- find certain situations where SVM is applied
- give concrete examples
- start out by talking about machine learning in general
- analogy w 2-d for equation of a line to generalize to multidimensional
- classification v regression
- explain how i made graph
- terminology in introduction (features, etc.)
- replace remaining photos with better quality
- rename accordingly

2 Introductory Notes

- optimal margin: gap between points on each side have smallest distance from separating line; this is what the SVM calculates
- linearly separable means there is a separating hyperplane where those values above are positive and below negative
- N x k data matrix and N x 1 vector Y that classifies the two species
- hyperplane given by
$$f(x) = w \cdot x + b = 0$$
- the vector w is the normal vector to the hyperplane (so wx1 - wx2 = 0)
- the distance from a point to the hyperplane is

$$\begin{aligned} D &= \frac{f(p)}{||w||} \\ &= ||(p - x_1)|| \cos\theta \\ &= \frac{w \cdot (p - x_1)}{||w||} \\ &= \frac{w \cdot p - w \cdot x_1}{||w||} \\ &= \frac{wp + b}{||w||} \\ &= \frac{f(p)}{||w||} \end{aligned}$$

- B^- is the max value of the minimums (those above the hyperplane)
- B^+ min value of the positives (below the hyperplane)
- supporting hyperplane: every point is on one side and you can't move it any further over otherwise that point will be on the wrong side
- geometric margin: the perpendicular distance between the two supporting hyperplanes

notes on 3094 book

- maximal margin classifier: separates two distinct classes
- support vector classifier / soft margin classifier: separates most data points according to a margin
- makes use of slack variables, epsilons that allow some points to be on the wrong side; if $\epsilon = 0$ it's on the right side of the margin, $\epsilon > 0$ wrong side of margin and $\epsilon > 1$ wrong side of hyperplane

- For C ¿ 0 no more than C observations can be on the wrong side of the hyperplane, because if an observation that sum of n violations to the margin, and so the margin will widen. Conversely, as C decreases, we become less tolerant of violations to the margin and so the margin narrows.

- support vectors: points that lie on the wrong side of the margin; these are the only observations that affect the classifier

- SKILEARN alphas are the same as his lambdas

-