ARTIFICIAL NEURAL NETWORKS

ARTIFICIAL NEURAL NETWORKS ARE A CLASS OF MACHINE LEARNING TECHNIQUES THAT WERE LOOSELY INSPIRED BY THE HUMAN NERVOUS SYSTEM -

"WERE" BECAUSE THIS WAS THE CASE WHEN THESE TECHNIQUES WERE FIRST PROPOSED IN THE 1950S - MODERN IMPLEMENTATIONS LOOK VERY MUCH LIKE OTHER STATISTICAL TECHNIQUES WE HAVE COVERED

"LOOSELY" BECAUSE THE COMPARISON WITH THE HUMAN NERVOUS SYSTEM HAS TENDED TO EVOKE IMPRESSIONS OF SELF-AWARE, ALMOST HUMAN COMPUTERS, THAT HAVE LITTLE TO DO WITH MOST OF THESE TECHNIQUES

ARTIFICIAL NEURAL NETWORKS ARE ANALOGOUS TO SUPPORT VECTOR MACHINES AND OTHER LEARNING TECHNIQUES -

INDEED, THE FIRST AND PROTOTYPICAL EXAMPLE OF AN ARTIFICIAL NEURAL NETWORK IS

A PERCEPTRON

WHICH IS BASICALLY A BINARY CLASSIFIER - SIMILAR TO A SUPPORT VECTOR MACHINE, BUT LESS SOPHISTICATED

ARTIFICIAL NEURAL NETWORKS HAVE HAD A HARD TIME LIVING UP TO SEVERAL DECADES OF HYPE -

BUT ANNS ARE NOW FINDING USE IN APPLICATIONS LIKE HANDWRITING ANALYSIS AND COMPUTER VISION

A PERCEPTRON IS A PROTOTYPICAL EXAMPLE OF AN ARTIFICIAL NEURAL NETWORK

PERCEPTRON

WHICH IS CONCEPTUALLY VERY CLOSELY RELATED TO A SUPPORT VECTOR MACHINE, WHICH WE ALREADY HAVE SPENT SOME TIME ON -

SUPPORT VECTOR MACHINES

A SUPPORT VECTOR MACHINE IS USED TO BUILD BINARY CLASSIFIERS

POINTS: EMAILS
CATEGORIES: SPAM OR HAM

(THIS MEANS THAT GIVEN A SET OF POINTS, A SUPPORT VECTOR MACHINE WILL [illegible] ASSIGN THOSE POINTS INTO 2 CATEGORIES

IN ADDITION, SUPPORT VECTOR MACHINES MAKE THEIR CLASSIFICATION DECISION ON THE BASIS OF A "LINEAR FUNCTION" OF THE POINT'S COORDINATES

IF A POINT IS $X = (X1, X2, X3, \ldots Xn)$

A LINEAR FUNCTION IS SOMETHING LIKE

$f(X) = aX1 + bX2 + cX3 + \ldots zXn$

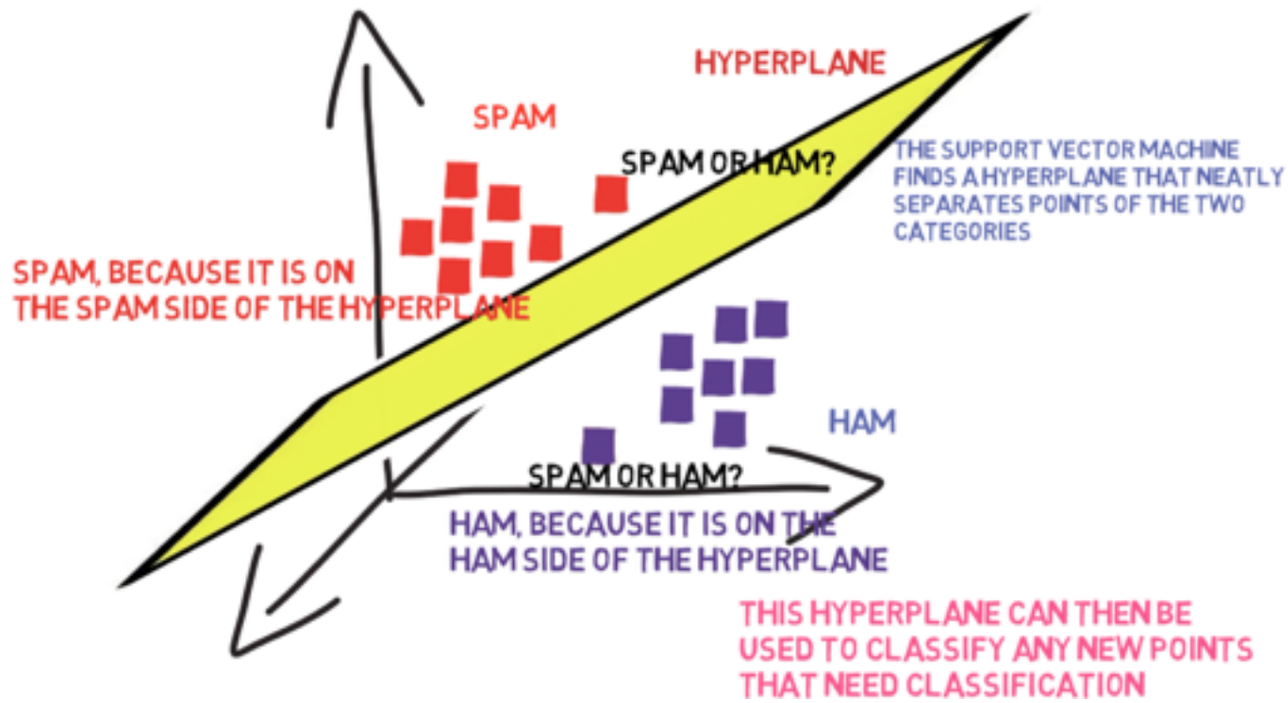
LASTLY, SVMS INVOLVE AN EXPLICIT TRAINING STAGE WHEN THE MODEL "LEARNS" FROM A SET OF TRAINING DATA

THE SUPPORT VECTOR MACHINE WILL RUN A TEST LIKE: IF $F(X) > 0$, EMAIL IS SPAM, ELSE EMAIL IS HAM

ALSO, SUPPORT VECTOR MACHINES DO NOT INVOLVE EXPLICIT ASSUMPTIONS ABOUT THE PROBABILITY DISTRIBUTIONS OF THE POINTS

(NAIVE BAYES CLASSIFIERS, FOR INSTANCE, ASSUME THAT THE DISTRIBUTIONS OF DIFFERENT FEATURES ARE INDEPENDENT)

"A SUPPORT VECTOR MACHINE IS A SUPERVISED MACHINE-LEARNING APPROACH USED TO BUILD LINEAR, NON-PROBABILISTIC BINARY CLASSIFIERS"



FIRST OFF - WHAT IS A HYPERPLANE?

IN A VECTOR SPACE OF N DIMENSIONS, A HYPERPLANE IS A GEOMETRIC SHAPE I.E. A SET OF POINTS - WITH (N-1) DIMENSIONS AND 0 THICKNESS IN ONE DIMENSION

THE EQUATION OF THE SET OF POINTS DEFINING THE HYPERPLANE IS ALWAYS "LINEAR"

$$Ax + By + Cz = D$$

IS THE EQUATION OF A HYPERPLANE IN 3D SPACE (I.E. A USUAL PLANE OF THE SORT WE JUST DREW)

ALL POINTS ON THE PLANE WILL SATISFY THIS EQUATION

ALL POINTS ON ONE SIDE OF THE PLANE WILL SATISFY THE CONDITION

$$Ax + By + Cz > D$$

AND ALL POINTS ON THE OTHER SIDE WILL SATISFY

$$Ax + By + Cz < D$$

THIS IS THE LINEAR EQUATION THAT THE SVM USES TO CLASSIFY POINTS - WHICH IS WHY THE SVM IS A LINEAR CLASSIFIER

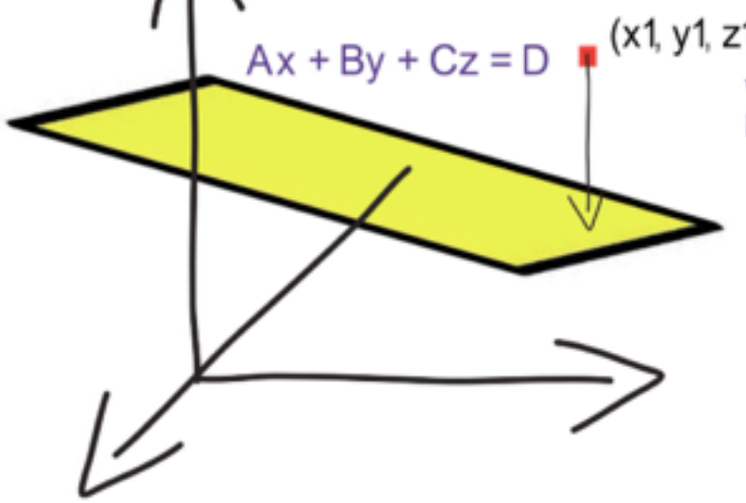
ALL POINTS ON ONE SIDE OF THE PLANE WILL SATISFY THE CONDITION

$$Ax + By + Cz > D$$

AND ALL POINTS ON THE OTHER SIDE WILL SATISFY

$$Ax + By + Cz < D$$

THIS IS THE LINEAR EQUATION THAT THE SVM USES TO CLASSIFY POINTS - WHICH IS WHY THE SVM IS A LINEAR CLASSIFIER



WHAT IS THE DISTANCE OF THE POINT FROM THE PLANE?

$$\frac{Ax1 + By1 + Cz1 - D}{[A^2 + B^2 + C^2]^{1/2}}$$

NOW COMING BACK TO OUR BASIC PROBLEM - HOW DOES THE SUPPORT VECTOR MACHINE FIND THE "BEST" HYPERPLANE TO SEPARATE THE 2 SETS OF POINTS?

INTUITIVELY, THE "BEST" HYPERPLANE IS ONE THAT:

MAXIMIZES SUM OF THE DISTANCES OF THE NEAREST POINTS ON EITHER SIDE — OBJECTIVE FUNCTION

CONSTRAINTS — (WHILE STILL MAKING SURE THAT ALL POINTS OF ONE TYPE ARE ON ONE SIDE OF THE PLANE, AND ALL POINTS OF THE OTHER ARE ON THE OTHER)

THIS IS SET UP BEAUTIFULLY AS AN OPTIMIZATION PROBLEM..

NOW COMING BACK TO OUR BASIC PROBLEM - HOW DOES THE SUPPORT VECTOR MACHINE FIND THE "BEST" HYPERPLANE TO SEPARATE THE 2 SETS OF POINTS?

THE SOLUTION IS CALLED
THE MAXIMUM MARGIN HYPERPLANE

INTUITIVELY, THE "BEST" HYPERPLANE IS ONE THAT:

MAXIMIZES SUM OF THE DISTANCES OF THE NEAREST POINTS ON EITHER SIDE

OBJECTIVE FUNCTION

CONSTRAINTS
(WHILE STILL MAKING SURE THAT ALL POINTS OF ONE TYPE ARE ON ONE SIDE OF THE PLANE AND ALL POINTS OF THE OTHER ARE ON THE OTHER)

THIS IS SET UP BEAUTIFULLY AS AN OPTIMIZATION PROBLEM

WE WON'T GO INTO THE DETAILS OF HOW EXACTLY THAT OPTIMIZATION PROBLEM IS FRAMED MATHEMATICALLY OR SOLVED -

BUT SUFFICE IT TO SAY THAT IT CAN BE CONVERTED INTO A FAIRLY STANDARD QUADRATIC PROGRAMMING PROBLEM FOR WHICH STANDARD SOLUTION TECHNIQUES EXIST

THE MAXIMUM MARGIN HYPERPLANE

IS FOUND - AND BTW THE "SUPPORT VECTORS" ARE SIMPLY THE "NEAREST POINTS" ON EACH SIDE - WHICH "SUPPORT" THE HYPERPLANE

NOW ALL OF THIS WAS ABOUT SUPPORT VECTOR MACHINES, BUT ALSO APPLIED TO

PERCEPTRONS