in this small lecture i will discuss a little bit looking under the hood of simple classification in a classification problem we're provided as our inputs a feature vector extracted from an input data point and typically in the training data now for visualization purposes i'm going to imagine that it's a two-dimensional feature vector and so we can plot it on the plane of the lecture slide so the first feature extracted from x is v1 of x and the second feature is v2 of x all the fees in this lecture are going to have x vector as their argument so this i'm going to leave it off the argument so v2 is v2 of x vector okay so here we have two dimensional diagram because our phi vector has two dimensions each training data point can be plotted in this plane so we have a space of possible training data points if we were imagining that our input x was an image and our v vector had many dimensions and extracted many features of the image there would be a point in a very high dimensional space for that image after feature extraction and that point will be labeled yes or no it's in the class likewise here each point may be labeled yes or no it's in the class and let's suppose we'll put a plus for those points labeled yes in the class and a minus for those points labeled no so i'm just going to plot a data set a potential data set okay so on this training data set the size of a minus shouldn't be varying um there we go on this training data set it's pretty clear that the pluses are sort of occurring down to the right and the minuses are occurring up to the left and in fact we might like to have a division somewhere along here right pardon me for the excitement of my dogs so when in fact our machine learning technique will learn something like this as a boundary it will adjust that boundary um by changing the angle of the slope through the origin to minimize whatever loss function you're minimizing in the case of hinge loss it will try to drive the margins of these points up above one and in its attempt to do that it's got only one thing it can adjust which is the weight vector and the weight vector is can also be plotted in this same real plane although weights are not values of features i'll still plot it in the same tray real plane and it's going to be this way how do i know it's going to be this way because the boundary is going to be a right angle to the weight vector so the magnitude of the weight vector does not change where the boundary is because we take the dot product between the weight vector and the vector represented by one of these plus or minus signs and we check the sign of that dot product and the sign of that dot product will tell you whether the angle between the two vectors is acute or obtuse whether it's less than 90 degrees or greater than 90 degrees and we can see all the plus vectors here are less than 90 degrees and all the minus vectors let's say one of these are greater than 90 degrees to the green so something like this will be our learned weight vector representing this decision boundary so this would give you a little bit of geometric interpretation of what's happening under the hood and we can clearly see just if i can draw it if we were to have a somewhat different problem right and also just for the for clarity it doesn't have to perfectly be separated there could be a plus lurking over there and the minuses and some minuses lurking over here in the plus we're still because of the quantity of data that we need to get right to the loss trade-offs of trying to get those outliers right won't be won't be favorable so we'll basically end up ignoring them and we will still end up with a decision boundary like this and again the weight vector will point from the origin at a right angle to that boundary so that's all simple enough but of course what must be occurring to anyone is boy these are simple concepts now to be noted we don't know how complicated phi 1 and phi 2 are maybe phi 1 consults an oracle such as my dad and says is this a cat and oracle says whether they think it's a cat and then this one consults my mom and whether my mom thinks it's a cat so i get two feature values that came from two extremely informed people about whether it's a cat and maybe it becomes a very easy problem humor aside we don't know what's inside of p1 and p2 to make to reduce the problem to this kind of linear boundary but to be noted we can only learn a hyperplane through the origin that's all this can do and that feels very limiting at first until you realize that it's a hyperplane through the origin in some abstract high dimensional space created by the feature extraction and that can be a very powerful thing now associated with this lecture when you finish this lecture you should watch the youtube video that gives you a little bit of dramatization of how raising the dimensionality can allow a hyperplane to do something that didn't look linear we'll also illustrate it here with one more example so the obvious kind of thing you might wonder about would be what if i wanted to say a circle what if this is 2 2 minus 2 minus 2 and what if i imagine i would like a decision boundary like this so most of my pluses are falling within this and the minuses are falling outside there may be exceptions can i get this well i've already told you no right all i can get is a weight vector

dotted with v it's going to represent a line in this space what if i used different features now i'm going to imagine using different features but plotting the result in on this graph and getting this result by simply taking psi 1 equal to phi 1 squared plus v2 squared psi 2 equal to 1. so this is a feature vector as well i've left off the arguments all of these have arguments x vector psi ones i two p one p two they all have arguments x vector i'm just saying instead of extracting whatever v one was i'll extract that and square it and also extract v2 and square it and then i'll add those together that'll be my first feature so it's just a more complicated feature extractor and my second feature is actually much simpler it just always returns one and now now then i will have a weight vector that's going to say let's see i'm going to have to give a weight of w2 equal to 4 and w1 equal to -1 and then i'll be taking the prediction on on my x vector i better put in psi that'll be how i left off the side i need to have the sign on this right and so you should see that this sign is positive when phi 1 squared plus v2 squared is less than 4. and that's precisely what i wanted here so if i use this feature extraction there is a weight vector that gives me this exact decision boundary so the decision boundaries are linear with respect to with respect to the features but very non-linear with respect to the features and the weights but very non-linear with respect to original data or other features okay okay the feature or the weight vectors are treated linearly but the but other concepts including the original data you can have a very non-linear plot all right so that's just a teaser on decision boundaries i've also mentioned earlier that by putting in features for arbitrary let's say your original data was a single real number and we were classifying that single real number we can put in features for arbitrary powers of that real number and get arbitrary polynomials computable by the dot product right so it'd be possible to say that we want only we want to say yes only to those real numbers that are close to a root of some arbitrary polynomial if we had the right features we could have a 10th degree polynomial we could say this real number is a yes answer if it's close to a root of this 10th degree polynomial by choosing the the right weight vector to represent that polynomial um all right so don't be fooled by the linearity the linearity however is what makes it the problem of minimizing squared loss convex so that gradient descent will find a global optimum we are linear in just the right things to make that possible so watch the video uh link to from the landing page for this lecture on youtube to see a visualization of some of what's going on down here unfortunately it has no words