Lecture 4 Notes

Introduction to Data Science IS360

1 Non-linear Transformation

1.1 The $\mathcal{Z}$ Space

All the linear models covered so far depend on the linear transformation:

$$\mathbf{w}^\top \cdot \mathbf{x}. \tag{1}$$

The hypothesis produced by this transformation assumes that a linear relationship exists in the feature space $\mathcal{X}$. However, this is not always the case, many real-world datasets exhibit complex, non-linear patterns that cannot be captured by a straight line or hyperplane.

Even when the input data exhibits non-linear qualities we might still want to use a linear model (for simplicity). We can imagine transforming the non-linear input space $\mathcal{X}$ into a linear $\mathcal{Z}$ *feature space*:

$$\Phi : \mathcal{X} \to \mathcal{Z}, \tag{2}$$

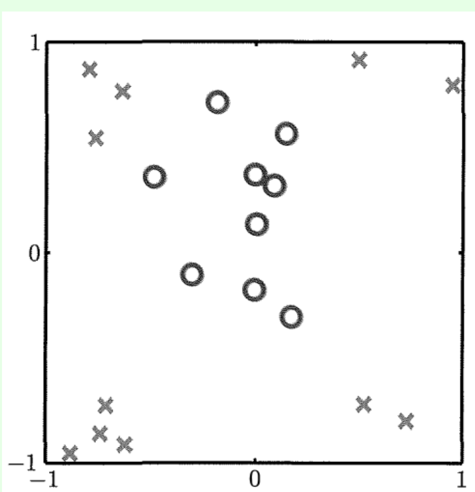
$$\mathbf{z} = \Phi(\mathbf{x}), \tag{3}$$

where Φ (Phi) is the *feature transform*.

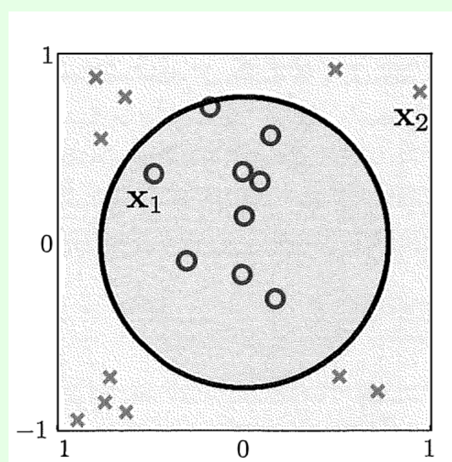
Note 1. We denote different quantities in $\mathcal{Z}$ using $\sim$ (tilde). For example, if $\mathcal{X} = \mathbb{R}^d$, then $\mathcal{X} = \mathbb{R}^{\tilde{d}}$, the weight vector $\mathbf{w}$ becomes $\tilde{\mathbf{w}}$ and so on. How exactly these quantities differ will depend on the transform itself Φ.

Non-linear Transform Example

Consider a classification task on following set of points:



(a) Original data $\mathbf{x}_n \in \mathcal{X}$.



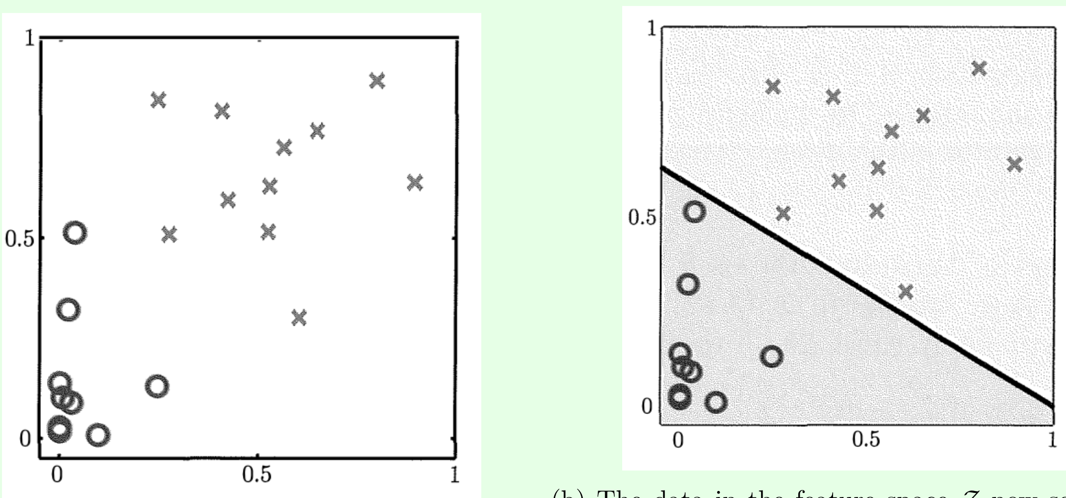
(b) In the original input space the data seems to be best separated using a circle.

We can imagine performing the following transform, with the goal of bringing our linear model ($\mathbf{w}^\top \cdot \mathbf{x}$) closer to the equation for the circle ($x_1^2 + x_2^2 = r^2$):

$$\mathbf{z} = \Phi(\mathbf{x}) = (1, x_1^2, x_2^2). \tag{4}$$

Notice how with this transformation multiple points in the input space may map to the same point in the feature space. For example, $(1, -2, -2)$ and $(1, 2, 2)$ in $\mathcal{X}$ will both map to the same point in $\mathcal{Z}$: $(1, 4, 4)$.

Applying this transformation to the data:



(a) Transformed data $\mathbf{z} = \Phi(\mathbf{x})$.

(b) The data in the feature space $\mathcal{Z}$ now seems to be linearly separable $\tilde{h}(\mathbf{z}) = \text{sign}(\tilde{\mathbf{w}}^\top \cdot \mathbf{z})$.

Based on the transformation we performed the final classifier in $\mathcal{X}$ would be:

$$h(\mathbf{x}) = \tilde{h}\left(\Phi(\mathbf{x})\right) = \text{sign}(\tilde{\mathbf{w}}^\top \cdot \mathbf{z}) \tag{5}$$

Common Non-linear Transformations
Polynomial Transformations

For an input vector $\mathbf{x} = [x_1, x_2, \cdots, x_n]$ a polynomial transformation of degree d includes all possible monomials up to degree d:

$$\Phi(\mathbf{x}) = \left\{ \prod_{i=1}^{n} x_i^{k_i} \,\middle|\, \sum_{i=1}^{n} k_i \leq d, k_i \geq 0 \right\} \tag{6}$$

For example a transformation with $d = 2$ would be:

$$\Phi(\mathbf{x}) = [x_1, x_2, \cdots, x_n, x_1^2, x_1 \cdot x_2, \cdots, x_n^2] \tag{7}$$

Radial Basis Function (RBF) Kernel

The RBF kernel maps data into an infinite-dimensional space using Gaussian functions centered at each data point:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp -\gamma ||\mathbf{x}_i - \mathbf{x}_j||^2, \tag{8}$$

where $||\cdot||^2$ denotes squared Euclidian distance, and $\gamma = \frac{1}{2\sigma^2}$ controls the width of the Gaussian function

Sigmoid Kernel

The sigmoid kernel uses the hyperbolic tangent function to project data into a non-linear space:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \tanh(\kappa \cdot \mathbf{x}_i^\top \cdot \mathbf{x}_j + c), \tag{9}$$

where κ (kappa) and c are kernel parameters that need to be specified.

1.2 Considerations