Best-Fit Conic Section Application

Code

In[∘]:=
```
data = {⋯} ; (* 60×2 matrix *)
ϕ[{x_, y_}] := {x^2, xy, y^2, x, y, 1};
```

Setting

Suppose we have m data points $\{x_1, x_2, \ldots x_m\}$ in $\mathbb{R}^2$.
In our working example we have $m = 60$.

A conic equation has the form

$$\mathcal{F}(x) = \mathcal{F}(x_1, x_2) = Ax_1^2 + Bx_1x_2 + Cx_2^2 + Dx_1 + Ex_2 + F$$

Which can be written as the following dot product:

$$\mathcal{F}(x) = \theta^{\mathsf{T}}\phi(x)$$

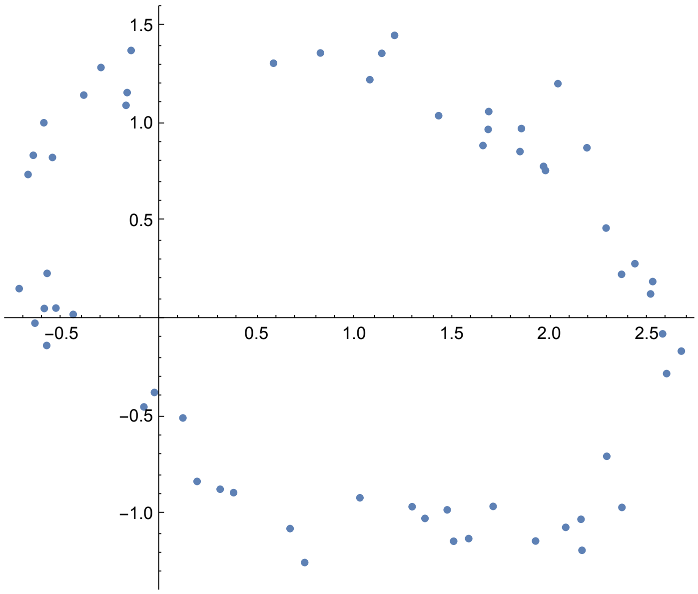
where

$$\theta = \begin{bmatrix} A \\ B \\ C \\ D \\ E \\ F \end{bmatrix}, \quad \phi(x) = \begin{bmatrix} x_1^2 \\ x_1x_2 \\ x_2^2 \\ x_1 \\ x_2 \\ 1 \end{bmatrix}$$

Goal:
The goal we set out to achieve is to find the function $\mathcal{F}$ so that the curve defined by $\mathcal{F}(x) = 0$ best approximates the data.

For example, our data may look like this:

In[∘]:=
```
ListPlot[data, AspectRatio → Automatic]
```

Out[∘]=


Finding the best-fit function $\mathcal{F}$ amounts to finding the vector θ of *parameters* such that the values $\{\theta^\top \phi(x_i) : i = 1 \ldots m\}$ are close to zero.

Loss

To make the notion of "best fit" more precise, we define the notion of "loss".
Given parameters θ and a set of data points $X = \{x_i\}_{i=1}^{m}$ we define the associated loss as

$$L(\theta, X) = \sum_{i=1}^{m} (\theta^\top \phi(x_i))^2.$$

The loss can qualitatively be described as the "total squared distance from data points to the curve." By way of analogy, consider the situation for linear regression:

If one aims to find a hyperplane $h^\top x + b = 0$ which best approximates some data, the associated loss for the parameters $\theta = (h, b)$ is given by:

$$L(\theta, X) = \sum_{i=1}^{m} (h^\top x + b)^2$$

which is the total squared distance to the hyperplane.
In fact, linear regression is exactly the task we are now faced with. Except the data points, originally in $\mathbb{R}^2$, have been mapped to points in $\mathbb{R}^6$ via the map $\phi(x)$.
In our case, for any chosen parameters θ the hyperplane has the equation

$$\theta^\top z = A z_1 + B z_2 + C z_3 + D z_4 + E z_5 + F z_6 = 0$$

and we want to find θ such that the losses $\theta^\top \phi(x_i)$ are minimized in total.

Notice there is no "b" term like in the linear regression - This means we're looking for a hyperplane $\theta^\top z = 0$ which passes through the origin in $\mathbb{R}^6$.

Solution

We start by expressing the loss as a matrix product.

Let Φ be an $m \times 6$ matrix where row i is given by $\phi(x_i)$.

$$\Phi = \begin{bmatrix} \phi(x_1) \\ \phi(x_2) \\ \vdots \\ \phi(x_m) \end{bmatrix}$$

In[∘]:=
```
Φ = Table[ϕ[x], {x, data}];
Print["Φ has size m×n where {m, n} = ", Dimensions[Φ]];
```

```
Φ has size m×n where {m, n} = {60, 6}
```

Then the loss is given by

$$\begin{aligned} L &= \| \Phi\theta \|^2 \\ &= \theta^\mathsf{T}\Phi^\mathsf{T}\Phi\theta \end{aligned}$$

Since the matrix $\Phi^\mathsf{T}\Phi$ is symmetric with real entries, there exists a square orthogonal matrix P and diagonal matrix D such that

$$\Phi^\mathsf{T}\Phi = PDP^\mathsf{T}$$

where $\Phi^\mathsf{T}\Phi$, D, P are all of size 6×6.

Then we can rewrite the loss as:

$$\begin{aligned} L &= (\theta^\mathsf{T}P)D(P^\mathsf{T}\theta) \\ &= (P^\mathsf{T}\theta)^\mathsf{T}D(P^\mathsf{T}\theta) \end{aligned}$$

We are almost done. There is one caveat about θ: Since we are finding the equation of a hyperplane, we are not looking for any one single θ per se. Any scalar of θ gives the same hyperplane equation so we are really looking for a *direction* that minimizes the loss. So lets restrict our search to the unit sphere $\|\theta\| = 1$.

Suppose the diagonal entries of D are $(\lambda_1, \lambda_2 \ldots \lambda_6)$. Then for any $y \in \mathbb{R}^6$,

$$y^\mathsf{T}Dy = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \lambda_3 y_3^2 + \ldots + \lambda_6 y_6^2$$

If $\|y\| = 1$ then the optimal$^*_{[1]}$ choice is to pick any j for which λ_j is minimal and set $y = e_j$.
For this choice, $y^\mathsf{T} D\, y = \lambda_j$ is the minimum possible loss.

We mention without proof that the following are equivalent:
(1) $\Phi^\mathsf{T}\Phi$ has linearly *dependent* columns
(2) D is singular
(3) $\lambda_i = 0$ for some i
(4) $N(\Phi^\mathsf{T}\Phi) \neq \{0\}$
(5) The data points lie *exactly* on the curve $\theta^\mathsf{T}\phi(x) = 0$

Thus the optimal$^*_{[1]}$ solution is given by $\theta = P\, e_j = P_{\text{col } j}$ where λ_j is the smallest eigenvalue of $\Phi^\mathsf{T}\Phi$.

In[∘]:=

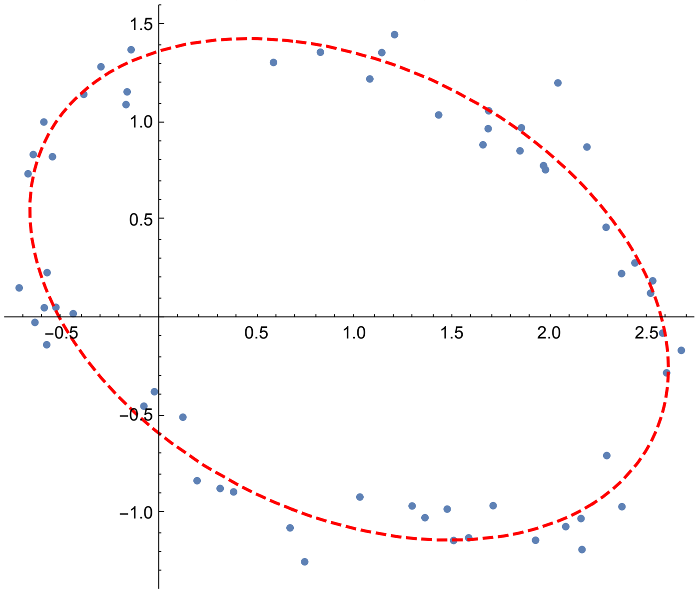
```
{eigenvalues, eigenvectors} = Eigensystem[Φᵀ.Φ];
(* sort the eigenvalues in increasing order, reorder the corresponding
   eigenvectors the same way, take the first one and scale it to unit length *)
θ = Normalize[First[eigenvectors[[Ordering[eigenvalues]]]]];
Print["θ = ", θ]
```

θ = {-0.2906, -0.2339, -0.4712, 0.5973, 0.3664, 0.3845}

```
Show[{ ListPlot[data, AspectRatio → Automatic],
        ContourPlot[θ.ϕ[{x, y}] == 0,  (* ⟵ this is the equation θᵀϕ(x) = 0  *)
           {x, -2, 3}, {y, -2, 3}, ContourStyle → Directive[Dashed, Red]]},
PlotLabel → Style[Evaluate[θ.ϕ[{x, y}] == 0], FontFamily → "Times"]]
```

Out[∘]=

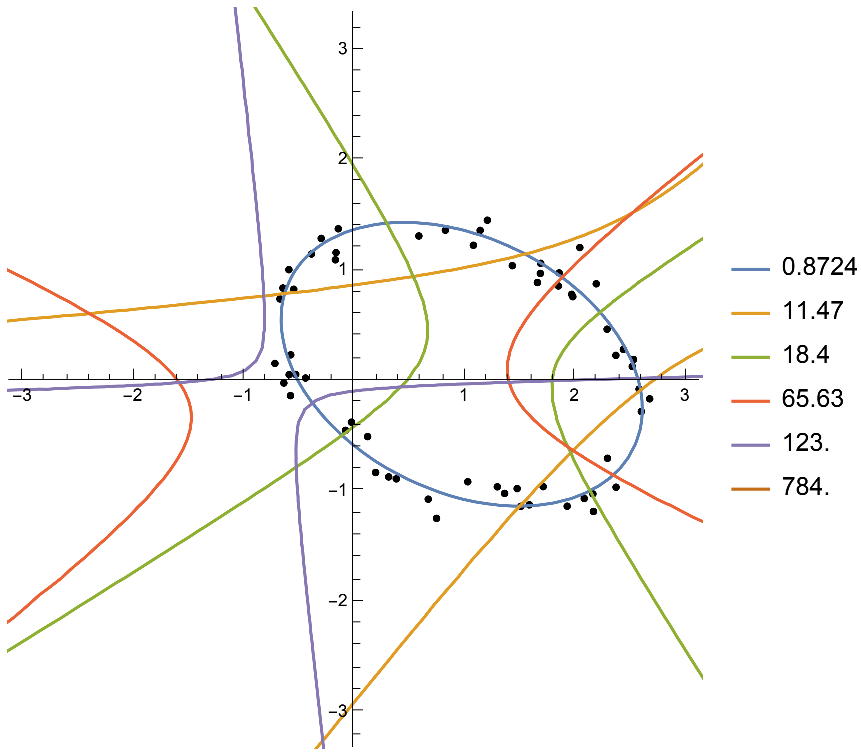
$-0.2906\,x^2 - 0.2339\,x\,y + 0.5973\,x - 0.4712\,y^2 + 0.3664\,y + 0.3845 = 0$



Associated with the other columns of P we obtain other, non-optimal curves.
Here we see that the other choices come nowhere near as close to approximating the data.
The eigenvalues for each choice of θ are shown to the right.

```
ord = Ordering[eigenvalues];
Show[{
    ListPlot[data, AspectRatio → Automatic, PlotStyle → Black],
    ContourPlot[
        Evaluate@Table[v.ϕ[{x, y}] == 0, {v, eigenvectors⟦ord⟧}],
        {x, -5, 5}, {y, -5, 5}, PlotLegends → eigenvalues⟦ord⟧]
    },
    PlotRange → {{-3, 3}, {-3, 3}}]
```

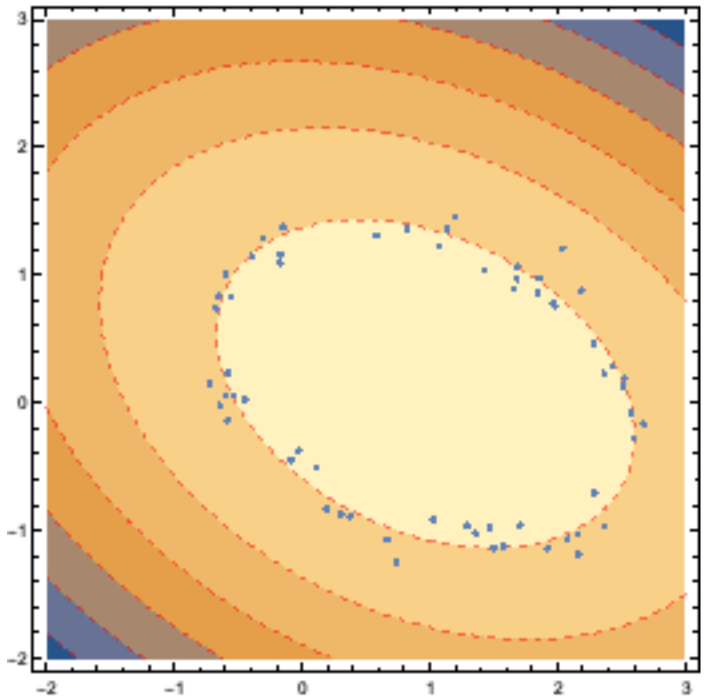
Out[∘]=



Here are a few other contours, which represent the curves where $\theta^{\mathsf{T}}\,\phi(x) = c$ for various values of c.

```
Show[{
    ContourPlot[θ.ϕ[{x, y}], {x, -2, 3}, {y, -2, 3}, ContourStyle → Directive[Dashed, Red]],
    ListPlot[data, AspectRatio → Automatic]
}]
```

Out[∘]=



Notes:

*[1]: We have that each $\lambda_i \geq 0$ since $\Phi^{\mathsf{T}}\,\Phi$ is positive semi-definite. This comes from the fact that $x^{\mathsf{T}}\,\Phi^{\mathsf{T}}\,\Phi\,x = ||\Phi\,x||^2$ which is non-negative for every x.

$$\Phi^{\mathsf{T}}\Phi v = \lambda v \Longrightarrow ||\Phi v||^2 = v^{\mathsf{T}}\Phi^{\mathsf{T}}\Phi v = \lambda ||v||^2 \geq 0$$