Introduction to Fenchel Duality for Machine Learning

Ryan M. Rifkin

Honda Research Institute USA, Inc.
Human Intention Understanding Group

2007

Motivation

- In machine learning (and in life), we frequently need to optimize a function.
- If the function is *convex*, then every local minimum is a global minimum. This is a very desirable property.
- If the function is *convex and differentiable everywhere*, then we can find the minimizer by setting the derivative to zero.
- If the function is *convex* but *not differentiable*, then we need a more advanced theory.
- Fenchel duality is one such theory.

Fenchel-Legendre Conjugate

Definition (Fenchel-Legendre conjugate)

Given a function $f : \mathbb{R}^n \rightarrow (-\infty, \infty]$, the *Fenchel-Legendre conjugate* is

$$f^*(z) = \sup_y \{y^t z - f(y)\}.$$

Fenchel-Legendre Conjugate, $\frac{1}{2}y^2$

$$
\begin{aligned}
f(y) &= \frac{1}{2}y^2 \\
f^*(z) &= \sup_y \{yz - f(y)\} \\
&= \sup_y \{yz - \frac{1}{2}y^2\}
\end{aligned}
$$

For any given z, the sup is attained when $y = z$:

$$
\begin{aligned}
f^*(z) &= z^2 - \frac{1}{2}z^2 \\
&= \frac{1}{2}z^2
\end{aligned}
$$

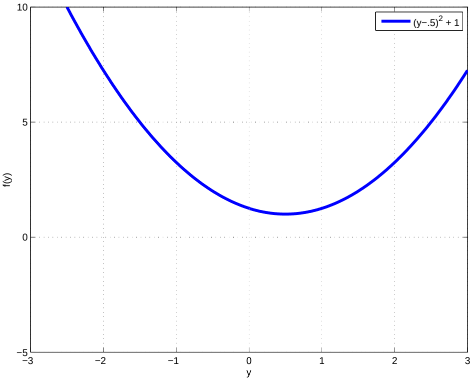
Fenchel-Legendre Conjugate, $\frac{1}{2}y^tAy$

$$
\begin{aligned}
f(y) &= \frac{1}{2}y^tAy \\
f^*(z) &= \sup_y \{y^tz - \frac{1}{2}y^tAy\} \\
&= \frac{1}{2}z^tA^{-1}z
\end{aligned}
$$

Fenchel-Legendre Conjugate, Additive Constant

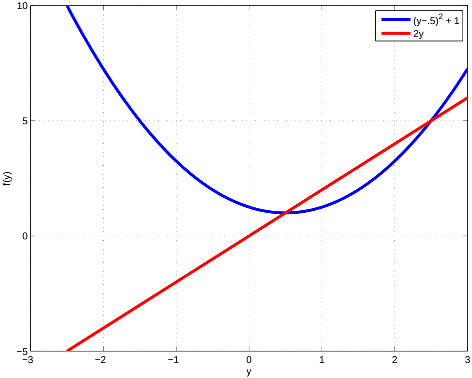
$$
\begin{aligned}
g(y) &= f(y) + c \\
g^*(z) &= \sup_y \{y^t z - g(y)\} \\
&= \sup_y \{y^t z - f(y) - c\} \\
&= f^*(z) - c
\end{aligned}
$$

Fenchel-Legendre Conjugate, Geometric View



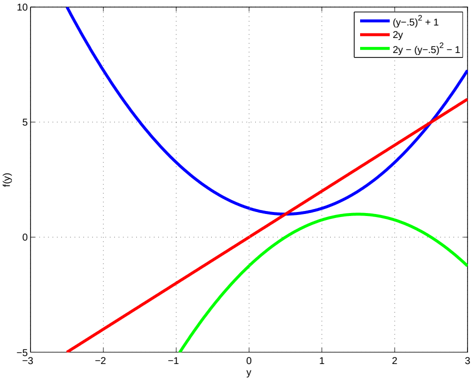
The function $f(y) = (y - .5)^2 + 1$...

Fenchel-Legendre Conjugate, Geometric View



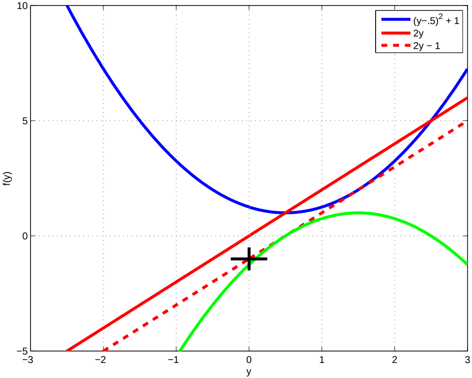
We add the line $f(y) = 2y$...

Fenchel-Legendre Conjugate, Geometric View



$f^*(2)$ is the sup of the difference ...

Fenchel-Legendre Conjugate, Geometric View



...and we can find the negative of the sup via a shift. The key point is that the sup occurs where the "slope" of f is z ...

Fenchel-Legendre Conjugate, Mini Homework

1. Let $g(y) = f(y) - a^t y$. Derive $g^*(z)$ (in terms of $f^*(z)$).
2. Let $g(y) = af(y)$ (assume $a > 0$). Derive $g^*(z)$.
3. Let $g(y) = f(a - y)$. Derive $g^*(z)$. (Hint: Your solution will be in terms of $f^*(-z)$.)
4. Let $f(y) = \frac{1}{2}(a - y)^2$. Derive $f^*(z)$ two different ways: first directly using the definition of the Fenchel-Legendre conjugate, and then by applying previously known identities.
5. Let $f(y) = |y|$. Derive $f^*(z)$. (Hint: This requires some effort. Stick with the Fenchel-Legendre conjugate definition, try some different values of z, and explore. Alternately, use the graphical method.)

Fenchel-Young (Part I)

$$\begin{aligned} f(y) + f^*(z) &= f(y) + \sup_{\hat{y}}\{\hat{y}^t z - f(\hat{y})\} \\ &\geq f(y) + y^t z - f(y) \\ &= y^t z \end{aligned}$$

Fenchel-Young, Convex Differentiable Case

Suppose f is convex and differentiable. Choose some (fixed) y.

$$\begin{aligned} f^*(\nabla f(y)) &= \sup_{\hat{y}} \{\hat{y}^t \nabla f(y) - f(\hat{y})\} \\ &= y^t \nabla f(y) - f(y), \end{aligned}$$

implying

$$z = \nabla f(y) \Longrightarrow f(y) + f^*(z) - y^t z = 0$$

Fenchel Duality Motivation

Frequently, we will be interested in minimizing the sum of two functions. We want to find

$$\inf_y \{f(y) + g(y)\}$$

Applying Fenchel-Young twice,

$$\begin{aligned}
f(y) + f^*(z) - y^t z &\geq 0 \\
g(y) + g^*(-z) - y^t(-z) &\geq 0 \\
f(y) + g(y) + f^*(z) + g^*(-z) &\geq 0 \\
\inf_y \{f(y) + g(y)\} + \inf_z \{f^*(z) + g^*(-z)\} &\geq 0.
\end{aligned}$$

If f and g are convex (plus minor technical conditions), then the infima are attained.

Theorem

Given convex functions f and g, under minor technical conditions,

$$\inf_{y,z} \{f(y) + g(y) + f^*(z) + g^*(-z)\} = 0,$$

at least one minimizer exists, and all minimizers y, z satisfy the complementarity equations:

$$\begin{aligned} f(y) - y^t z + f^*(z) &= 0 \\ g(y) + y^t z + g^*(-z) &= 0. \end{aligned}$$

The Differentiable Case

Suppose f and g are both convex and differentiable. Then the complementarity conditions become

$$\begin{aligned} z &= \nabla f(y) \\ -z &= \nabla g(y), \end{aligned}$$

which combine into

$$\nabla f(y) + \nabla g(y) = 0,$$

the well-known optimality condition for $f + g$.

More information

These slides, as well as more slides and other information on Fenchel Duality and machine learning, are available at http://middleangle.com/rif/derifatives/fenchel