Incentives and Machine Learning (CSC 482A/581A)
Lectures 4 and 5

Nishant Mehta

1 Information elicitation

Suppose that you would like to obtain a weather forecast for tomorrow. The possibilities are rain or sun. Not knowing much about how to forecast the weather, you decide to recruit an expert. The expert has a probabilistic belief about tomorrow's weather in the form of the probability of rain. Thus, from the expert's perspective, the outcome to be forecast is drawn from a Bernoulli distribution with success probability p. Unfortunately, the expert is selfish and, without any monetary compensation, she cannot be trusted to honestly report the weather. You therefore wonder:

> Is there a mechanism I could design which, given the expert's reported probability of rain and the actual outcome, pays the expert in such a way that she maximizes her expected payment by reporting her true belief p?

The above question is at the heart of *information elicitation*. In information elicitation, there are some experts that possess private information, and the mechanism designer would like to design a mechanism that incentivizes the agents to reveal their private information. What are experts? Well, they could be actual experts (like a meteorologist in the weather forecasting example), a human worker with some better-than-nothing-but-not-necessarily-great belief, or even AI agents. Each expert has some probabilistic belief about some outcome, such as the weather on a future day or which country will win the gold medal in men's downhill skiing in the 2028 Winter Olympics.

Let us now formalize a forecasting game. First, we extend the previous setting to a more general, "multi-class" setting. Suppose that there is a finite set $\mathcal{Y}$ of K outcomes; we often equate Y with the set $\{1, 2, \ldots, K\}$. An expert's belief p is now a probability vector in the probability simplex over K outcomes, denoted as $\Delta(\mathcal{Y})$. We wish for a mechanism that incentivizes an expert to reporting truthfully. Concretely, consider the following game protocol:

1. Expert reports probability forecast $r \in \Delta(\mathcal{Y})$.
2. Mechanism observes outcome $y \in \mathcal{Y}$.
3. Mechanism pays expert $S(r, y)$, where S is scoring rule $S \colon \Delta(\mathcal{Y}) \times \mathcal{Y} \to \mathbb{R}$.

The question from before is how the mechanism can choose S so that the expert is incentivized to report truthfully (by setting $r = p$). To engage with this question, we must make assumptions about the expert's behavior. We assume that an expert tries to maximize her expected payment, where she computes expectations according to her belief p.

So, we seek a scoring rule such that for any belief $p \in \Delta(\mathcal{Y})$ and any report $r \in \Delta(\mathcal{Y})$ with $r \neq P$, it holds that

$$\mathsf{E}_{Y \sim p}[S(p, Y)] \geq \mathsf{E}_{Y \sim p}[S(r, Y)].$$

Even better, we hope for strict inequality, so that an expert is strictly better off by reporting truthfully.

Since we so often wish to refer to the expected score, let's adopt the notation

$$S(r, p) = \mathsf{E}_{Y \sim p}\left[S(r, Y)\right].$$

Definition 1. Let $S\colon \Delta(\mathcal{Y}) \times \mathcal{Y}$ be a scoring rule. We say that S is *proper* if, for all $r \in \Delta(\mathcal{Y})$ with $r \neq p$,

$$S(p, p) \geq S(r, p).$$

Moreover, S is *strictly proper* if the inequality is strict.

1.1 Examples

There are many examples of proper scoring rules. Here, we just briefly discuss a few of them. The first example is the *quadratic score*, also known as the *Brier score*, defined as

$$S(r, y) = 1 - \sum_{j=1}^{K} (r_j - (\mathbf{e}_y)_j)^2,$$

where $\mathbf{e}_y$ is the y^{th} standard basis vector.

The quadratic score is strictly proper. Note that there is some flexibility in defining this score. The shape of the scoring rule would be the same if we dropped the additive 1. If we changed the additive 1 to 2, then the score would always be nonnegative (which may be desirable to an agent who never wishes for their payment to be negative). Also, note the similarity to the quadratic loss, $\ell(r, y) = \sum_{j=1}^{K} (r_j - (\mathbf{e}_y)_j)^2$.

The case of $K = 2$ is often treated in a special way. Letting the outcome space by $\mathcal{Y} = \{1, 0\}$, we use a single-dimension parameterization so that $p, r \in [0, 1]$, and then define $S(r, y) = 1 - (r - y)^2$.

The second example of a proper scoring rule is the log score (which also is strictly proper), defined as

$$S(r, y) = \log r_y.$$

Note that the log score is nothing more than the negation of the log loss. In the binary case, we often use the one-dimensional parameterization $S(r, y) = y \log r_y + (1 - y) \log r_{1-y}$, which is just the negation of the binary cross-entropy loss.

The log score has many beneficial properties. For example, when an outcome y occurs, the log score only pays a forecaster according to the probability the forecaster assigned to y. One not-so-great property is that the log score is unbounded below, which means that no amount of positive additive shift will ensure that the forecaster always receives a nonnegative payment.

We will soon see a characterization of the class of all proper scoring rules. Prior to that, we need to develop some basic concepts related to convexity.

2 Convexity background

Let $f\colon \mathbb{R}^d \to \mathbb{R}$ be a differentiable, convex function. Then for all $x, y \in \mathbb{R}^d$,

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle. \tag{1}$$

If f is also strictly convex function, then the inequality is strict whenever $x \neq y$.

In words, the inequality states that the function f is lower bounded by its first order Taylor approximation around x. We therefore will call this inequality the "first-order inequality". See the diagram below.

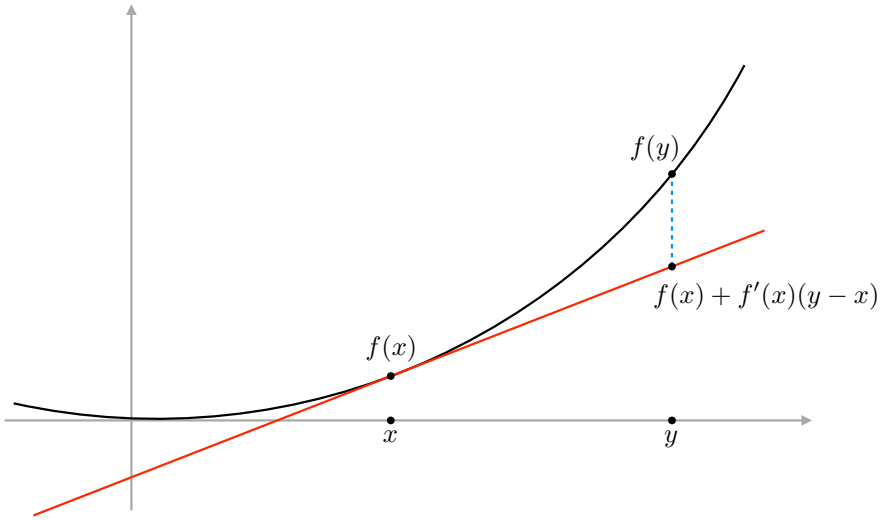

Figure 1: First-order inequality. If f is convex, any tangent line is never above f. This generalizes to the case of multi-dimensional domains, in which case the tangent line becomes a tangent hyperplane.

3 Characterization of proper scoring rules

How many proper scoring rules are there? There are infinitely many. The following lemma shows how, starting from a convex function, one can generate a corresponding proper scoring rule. The lemma assumes differentiability, but this can be relaxed (see the reddish box after the proof).

Lemma 1. *Let $G\colon \Delta(\mathcal{Y}) \to \mathbb{R}$ be a differentiable, convex function. then the scoring rule S defined as*

$$S(r, p) = G(r) + \langle \nabla G(r), p - r \rangle. \tag{2}$$

is proper. Moreover, If G is differentiable and strictly convex, then S is strictly proper.

Proof. To show that S is proper, we must show that, it holds that $S(p, p) \geq S(r, p)$ for all $r \in \Delta(\mathcal{Y})$. The proof easily follows by rewriting $S(p, p)$ and $S(r, p)$ using (2) and then appealing to the first-order inequality (1). Indeed, $S(p, p) = G(p)$ and $S(r, p) = G(r) + \langle \nabla G(r), p - r \rangle$. Hence, the desired inequality $S(p, p) \geq S(r, p)$ is equivalent to

$$G(p) \geq G(r) + \langle \nabla G(r), p - r \rangle. \tag{3}$$

Finally, note that (3) is true since G is convex.

If G is strictly convex, then we use (1) with strict inequality (for $r \neq P$) to get strict properness. □

A generalization. Lemma 1 can be generalized to the case of non-differentiable G. In that case, there can be reports at which $\nabla G(r)$ does not exist; as a simple example, consider $G(r) = |r - \frac{1}{2}|$. Yet, there is always at least one *subgradient* of G at r. We say that $g \in \mathbb{R}^K$ is a subgradient of G at r if, for all $p \in \Delta(\mathcal{Y})$,

$$G(p) \geq G(r) + \langle g, p - r \rangle;$$

this inequality is precisely what is needed in the proof. Hence, for not-necessarily-differentiable (but convex) G, we may define S as in (2) but, for each r, replace $\nabla G(r)$ with *any* subgradient of G at r. Finally, we mention that if G is differentiable at r, then $\nabla G(r)$ is the unique subgradient of G at r. Indeed, since $r \mapsto G(r) + \langle \nabla G(r), p - r \rangle$ is tangent to G at r, replacing $\nabla G(r)$ with any other vector must make the resulting tangent plane go above G somewhere in a neighborhood of r.

It is nice that we have a way to generate proper scoring rules; however, a natural question is whether there are some proper scoring rules that cannot be generated in the above way. That is, for any proper scoring rule, can we identify a corresponding convex function G that generates it? As the following lemma shows, the answer is yes, so this is a two-way street.

Lemma 2. *Let S be a differentiable proper scoring rule. Then there exists a differentiable, convex function $G\colon \Delta(\mathcal{Y}) \to \mathbb{R}$ such that S can be written as*

$$S(r, p) = G(r) + \langle \nabla G(r), p - r \rangle.$$

Moreover, if S is strictly proper, then G is strictly convex.

Proof. This direction is the more difficult one.

For any r, define the mapping $S_r(p)\colon p \mapsto S(r, p)$. Observe that S_r is a linear function of p since

$$S_r(p) = \mathsf{E}_{Y \sim P}\left[S(r, Y)\right] = \sum_{j=1}^{K} p_j S(r, j).$$

Therefore, $S(r, p)$ can be expressed as

$$S(r, p) = a + \langle v, p \rangle \tag{4}$$

for some $a \in \mathbb{R}$ and $v \in \mathbb{R}^K$ that can depend on r (but not on p).

Next, define the function $G(r) = S(r, r)$. Applying (4), we have

$$G(r) = S(r, r) = a + \langle v, r \rangle,$$

and so

$$a = G(r) - \langle v, r \rangle.$$

Plugging this expression for a into (4) gives

$$S(r, p) = G(r) + \langle v, p - r \rangle.$$

Now, since $S(r, p)$ exists, we know there must exist some v (which we recall can depend on r) satisfying the above equality. Are there any constraints that v must satisfy? So far, we have not

used the fact that S is proper. Let us use this fact. Properness implies that $S(p,p) \geq S(r,p)$, which using $S(p,p) = S_p(p) = G(p)$ and $S(r,p) = S_r(p)$, we rewrite as

$$G(p) \geq G(r) + \langle v, p - r \rangle.$$

We would be done if G were convex and if $v = \nabla G(r)$. We now show that G is convex and then, using this fact, show that indeed v must be equal to $\nabla G(r)$.

To see why G is convex, observe that $G(q) = S(q,q) = \max_{r \in \Delta(\mathcal{Y})} S_r(q)$. Every function S_r is convex (since it is linear in q), and the pointwise maximum of convex functions is also convex.[1] Hence, G is indeed convex.

Now, since G is convex, the first-order inequality (1) implies that

$$G(p) \geq G(r) + \langle \nabla G(r), p - r \rangle. \tag{5}$$

Observe that the choice $v = \nabla G(r)$ is feasible in light of (5). Is there any other choice for v? There cannot be, for $p \mapsto G(r) + \langle \nabla G(r), p - r \rangle$ is the unique tangent plane that touches G at r; any other choice (in particular, swapping out $\nabla G(r)$ for any other vector) would give a function that somewhere is above G (this is best visualized via a picture of a convex function from $\mathbb{R}$ to $\mathbb{R}$). Hence, for given r, the vector v uniquely takes the value $\nabla G(r)$. □

A generalization. Similar to the previous "generalization" comment, Lemma 2 can be generalized to the case where S is not differentiable. In this case, at each r, rather than using $\nabla G(r)$ (which can fail to exist), there exists a suitable choice of subgradient of G at r such that the lemma still holds. We avoid further elaboration in these notes.

4 Examples

Recall the quadratic score, defined as

$$S(r,y) = 1 - \sum_{j=1}^{K} (r_j - (\mathbf{e}_y)_j)^2$$

What convex function G corresponds to the quadratic scoring rule? Well, writing $G(p) = S(p,p)$ gives

$$\begin{aligned}
G(p) &= \mathsf{E}_{Y \sim p} \left[1 - \sum_{j=1}^{K} (p_j - (\mathbf{e}_Y)_j)^2 \right] \\
&= 1 - \|p\|_2^2 + \mathsf{E}_{Y \sim p} \left[\sum_{j=1}^{K} (2p_j (\mathbf{e}_Y)_j - (\mathbf{e}_Y)_j \right] \\
&= 1 - \|p\|_2^2 + \mathsf{E}_{Y \sim p} \left[\sum_{j=1}^{K} 2p_j^2 - p_j \right] \\
&= \|p\|_2^2.
\end{aligned}$$

[1] Showing that the pointwise maximum of convex functions is convex is a basic exercise.